

G., K., Okamura, S., Shimizu, S., Takeuchi, C., Wada, I., Watanabe, A., Yamada, M., Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76169

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-490 <KAN>

A:Cross-references: EMBL: D50914; GR: A600133; NID: g1553477; PIRN: SAA18428; i. EMBL: D50914
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 14.0%; Score 395; DB 2; Length 490;

Best Local Similarity: 28.0%; Pred. N°: 2; L: 20;

Matches 144; Conservative 75; Mismatches 211; Indels 84; Gaps 21;

QY 19 EVQGGPWTGCVLPGNPGMHITIGTKYNNHFWFGIALHHSFTFKNGEYVYYSKYLPSD 77
DQ 19 EVQGGPWTGCVLPGNPGMHITIGTKYNNHFWFGIALHHSFTFKNGEYVYYSKYLPSD 77
DQ 37 DVEGSIPLDGLGTLRNSPSLLEISLPLKHPFDMKVIATKFPGLGKRVHFSKVRVTO 95
QY 78 TYNCTNEANRIVV-SEGTIMAYPDCKNIKAKAFSYLSHTIPEFTDNCNLINMKAGUYY 136
DQ 78 TYNCTNEANRIVV-SEGTIMAYPDCKNIKAKAFSYLSHTIPEFTDNCNLINMKAGUYY 136
DQ 97 GYVEQKAKMYRCVGVGSPAGWALKTFINIFNTANI-----NITYWQDRL 145
QY 137 ATSETNFIKIDPOTLFTLKVDSKYVAVNIATS-HPHYDSAGNINLMGTSIVDK3PTK 195
DQ 137 ATSETNFIKIDPOTLFTLKVDSKYVAVNIATS-HPHYDSAGNINLMGTSIVDK3PTK 195
DQ 147 ALWEGGQPHLEPSNLATIGLDDGLAEGOPLSAHPRIDPA-----STFDGSGOPC 198
QY 196 YVIFKIPSPVPEKSKCFKHLVCSIPSRLQD-----SYHSFGITENYIVP 247
DQ 196 YVIFKIPSPVPEKSKCFKHLVCSIPSRLQD-----SYHSFGITENYIVP 247
DQ 199 YVIFKIPSPVPEKSKCFKHLVCSIPSRLQD-----SYHSFGITENYIVP 249
QY 248 IEQPKKIDVILATAYIRGVNWA-SCLSPKHKDKTWFHFVDKRTKKEVSTKFTDALVLY 306
DQ 248 IEQPKKIDVILATAYIRGVNWA-SCLSPKHKDKTWFHFVDKRTKKEVSTKFTDALVLY 306
DQ 250 LQNNVTN----GLPYLEGLRGACGCVQFHPDPAQIILVPRGGE--IKRIPVQACFPV 303
QY 307 HHINAYEDGHWVDIVAYRNSLYDMFYLLKKLDKDEPVANKLTSIP-TCKRFPVPLQYD 365
DQ 307 HHINAYEDGHWVDIVAYRNSLYDMFYLLKKLDKDEPVANKLTSIP-TCKRFPVPLQYD 365
DQ 304 HHANAFENGKTIIDSLCY--NSLPQV---DITDGFSTNFDNLDGQILWRFTID---- 353
QY 366 KLAEGVSNVRLPISAVKKEGSIYQPELCEGIELPEVAVYDNGKKYKVVY--ATE 423
DQ 366 KLAEGVSNVRLPISAVKKEGSIYQPELCEGIELPEVAVYDNGKKYKVVY--ATE 423
DQ 354 -----PAAATVEKOLMVSRG-----EFPVVPQGVGPRYRVYWGAAH 392
QY 424 VQWSHVPTK-TAKLNVT-KEVLHWGEDHGWSEPIFVSEDAFHEDEWVITVYVVSSE 481
DQ 424 VQWSHVPTK-TAKLNVT-KEVLHWGEDHGWSEPIFVSEDAFHEDEWVITVYVVSSE 481
DQ 393 HSTGNAPIQATIKVLESSEHFLWSPFALHGFACGPIFVSPFQCVALLGQWIL-GLIYKAD 454
QY 482 NKAPFLLIDAKTEKELGRATVRYVMIL--DLHC 513
DQ 482 NKAPFLLIDAKTEKELGRATVRYVMIL--DLHC 513
DQ 452 LHRSEVLVLDAGDITATATATATKELKHHTIPYTLHC 485

RESULT 5

AE2341

lignostilbene-alpha,beta-dioxygenase [imported] - Nostoc sp. (strain PCC 7120)

A:Species: Nostoc sp.

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120

C:Date: 14-Nov-2001 #sequence_revision 14 last:2001 #text_change 40-Jan-2002

C:Accession: AE2341

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriuchida

DNA Res. 3, 205-215, 2001

A:Reference number: AR1807; MUID:2159285; PMID:11759840

A:Accession: AE2341

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-497 <KUR>

A:Cross-references: GR:BA000010; PIRN:BAH75983; i. PIRN:g1733419; GSPNR:GN00179

C:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4284

Query Match 13.3%; Score 376.5; DB 2; Length 497;

Best Local Similarity: 25.6%; Pred. No. 4.5e-19;

Matches 133; Conservative 93; Mismatches 211; Indels 83; Gaps 21;

QY 19 EVQGGPWTGCVLPGNPGMHITIGTKYNNHFWFGIALHHSFTFKNGEYVYYSKYLPSD 78
DQ 19 EVQGGPWTGCVLPGNPGMHITIGTKYNNHFWFGIALHHSFTFKNGEYVYYSKYLPSD 78
DQ 38 DVEGSIPLDGLGTLRNSPSLLEISLPLKHPFDMKVIATKFPGLGKRVHFSKVRVTO 97
QY 79 TYNCTNEANRIVV-SEGTIMAYPDCKNIKAKAFSYLSHTIPEFTDNCNLINMKAGUYY 137
DQ 79 TYNCTNEANRIVV-SEGTIMAYPDCKNIKAKAFSYLSHTIPEFTDNCNLINMKAGUYY 137
DQ 99 YLAEKNAGKILHFGVFTQPPGWLANIFDFEKLNIANT-----NVIWGEKLLA 147
QY 148 LSLNIPFELTLELVYKRYVAVNLA-TSHHNYSA-----EKLHMTSTVAVK 191
DQ 148 LSLNIPFELTLELVYKRYVAVNLA-TSHHNYSA-----EKLHMTSTVAVK 191
DQ 148 LSLNIPFELTLELVYKRYVAVNLA-TSHHNYSA-----EKLHMTSTVAVK 191
QY 192 GPTKYVILFKIPSSVPEKSKCFKHLVCSIPSRLQD-----SYHSFGITENYIVP 251
DQ 192 GPTKYVILFKIPSSVPEKSKCFKHLVCSIPSRLQD-----SYHSFGITENYIVP 251
DQ 208 LSTTITITEL--NVAGEVWRKHA-HHVTGFC-----FIHDFVILPHNICIFQNP 253
QY 252 FRIGFVELATAYIRGVNWA-SCLSPKHKDKTWFHFVDKRTKKEVSTKFTDALVLY 306
DQ 252 FRIGFVELATAYIRGVNWA-SCLSPKHKDKTWFHFVDKRTKKEVSTKFTDALVLY 306
DQ 254 VSPNIPFALC-LRGA--GEGIQVQNPQPTKIVIPFPQSGAKETAT-IFVSGSIFPH 409
QY 309 INAYEFCHVVFIVAYFENSLYMEYLLKLLKDF-EVNNKITSTPTCKFFVVPVLDVKT 467
DQ 309 INAYEFCHVVFIVAYFENSLYMEYLLKLLKDF-EVNNKITSTPTCKFFVVPVLDVKT 467
DQ 310 VNAFVGEELVDSLCY--DSLPEV---EPESDFQVDFEATAPQOLMKPYLNL---KD 360
QY 368 AFVGSNIVKILPSTATAYKFGSTGYCTPEILGCEIFLPPVNYVNYKYVYVYVATEVQWS 427
DQ 368 AFVGSNIVKILPSTATAYKFGSTGYCTPEILGCEIFLPPVNYVNYKYVYVYVATEVQWS 427
DQ 361 GTVQKKLT-----ESPC--CEPPAIHPNVNSVYLYMSAAH-- 396
QY 428 PVPKIKAKI NVGLKEVLRHWRGLHW-----PSEKIFVLSIAPKFFPQVVI 480
DQ 428 PVPKIKAKI NVGLKEVLRHWRGLHW-----PSEKIFVLSIAPKFFPQVVI 480
DQ 397 -AATGNAPTQALLKIDILPSGKQIWSVAPPGFTGPIFVPPVPSFKFQDQAWIATVYDAA 455
QY 481 PNKAPFLLIDAKTEKELGRATVRYVMIL--DLHGMFIPQ 518
DQ 481 PNKAPFLLIDAKTEKELGRATVRYVMIL--DLHGMFIPQ 518
DQ 456 HHRSD-VVILDASDPTKGAIALRLHQLHVPVGLHGNFTPO 494

RESULT 6

TS1734

hexamethin cleavage enzyme [imported] Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 14 last:2000 #text_change 19-Aug-2000

C:Accession: TS1734

R:Reilly, S.O.; Burnett, E.C.; Besikova, R.; Hancock, J.T.

J. Exp. Bot. 49, 1893-1894, 1998

A:Title: Cloning of a wheat responsive cDNA from an Arabidopsis thaliana suspension cu

A:Reference number: 224454

A:Accession: TS1734

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-538 <NET>

A:Cross-references: EMBL:AJ005M13; PIRN:CAA06712.1

A:Experimental source: cultivar Landsberg erecta

C:Genetics:

A:Gene: ncl

Query Match 11.0%; Score 311; DB 2; Length 538;

Best Local Similarity: 23.8%; Pred. No. 2.3e-14;

Matches 136; Conservative 78; Mismatches 213; Indels 44; Gaps 25;

QY 9 KEHPEPTK-AEVQGLPTWLGVLPLNGPGMHITIGTKYNNHFWFGIALHHSFTFK 62
DQ 9 KEHPEPTK-AEVQGLPTWLGVLPLNGPGMHITIGTKYNNHFWFGIALHHSFTFK 62
DQ 56 RDETP-PVKDLPVGVFLPECLNGEFVVRGP-----NPKFDVAGYHWFDCGMHGVRI 108
QY 63 KMTVEVYPSKYLRSPTNYNMTNEANPIVVSFQTMAYFDPCKNIFAKAFSYLSHTIPEFTD 122
DQ 63 KMTVEVYPSKYLRSPTNYNMTNEANPIVVSFQTMAYFDPCKNIFAKAFSYLSHTIPEFTD 122
DQ 109 KMTKATVVSRY-----VKTSLKQEEF-----FGAAKFMKIGDLKGFPG 147

QY 124 NCLINL---MKTGDDYATSETN-----FIRKIDPOTLETLD 156
 DB 148 LLMNVVOOLTKILNLTNGTANTALVYHCKLALQEAHPVVIKVLGDGLQTLG 207
 QY 157 KVDYSKYVAVNLATSHPHYS-AGNINMCTISVDKGRKYVLFKIPSSVPEKKEKSCF 215
 DB 208 LIDYUKRLTHSF-TAHPKVDPTGEMTEFGYS---HTPPYLTYRVS-----KDG1 254
 QY 216 KHELVVCSIPSKSLQNSYHSHFCITENYIVFIDQFPL---DIVKLATAYINGVWASC 272
 DB 255 MHDVPPTIS-----EPIMHDFAITETYAIPMLDMLHHPKEMVK-----EKKM1 300
 QY 273 LSFHKEDKTWFHVDRAKTKKEVSTK-FYTDALVLYHHINAYEEDGHVFDIVAYRNSLY 331
 DB 301 YSFDPYTKARQGVLPYAKUDELMPWFELNCFIFINANAEDEEVL-ITCRLENPDL 359
 QY 332 DMFYLLKLDKDFEVNNKLTSIPTCKRNVVPLQYDKDAEVSNLVKLPTSATAVEKDGSI 391
 DB 360 DMVSGKVKLEKLENGNELYEM-----RF-----NMKTGSA 389
 QY 392 YCQPELICEGLPRVAVNYGKKYKYVYATEVQWSPVPTKIAKLVQV-----440
 DB 390 -SQKLSASAVDFPRNRYTGKKQRYVYGTILUSTAKVGTIKPDLHAEAEKGMLEV 448
 QY 441 ----KEVLHMGEDHCWSEPIFVSPDAREDEGVLTICVVSSEPNKAPFLILLDAKTEK 496
 DB 449 GGNIKGIYDGEGR-YGSEALYVPEFA-BEIDGCLIFFVHDENIGKS-CVTIVIDAKIMS 505
 QY 497 ELGRATVNV--FMHLDLHGMFIPNDLCAET 525
 DB 506 AEPVAVVELPHRVPGYGFHALFVTEQLEQQT 536

RESULT 7

T49193

neoxanthin cleavage enzyme ncl - Arabidopsis thaliana

N:Alternate names: protein MAA21.150

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Jun-2000 #sequence_revision: 02-Jun-2000 #text_change: 02-Jun-2000

C:Accession: T49193

K:Requester: M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Newes, H.W.; Rudd, S.; Lemcke, K.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z25018

A:Accession: T49193

A>Status: preliminary

A:Molecule type: DNA

A:Positives: 1-538 cPFS

A:Cross-references: EMBL:AL163a18; GSPDR:GN06061; ATSP:MAA21 150

A:Experimental source: cultivar Columbia; BAC clone MAA21

C:Genetics:

A:Gene: ATSP:MAA21.150

A:Map position: 3

A:Insertion: 72/3 56/2, 151/2, 158/3, 172/1, 193/1, 202/1, 207/1, 377/2, 411/2, 43

Query Match

Best Local Similarity 11.0%; Score 310; DB 2; Length 538

Matches 145; Conservative 70; Mismatches 213; Indels 144; Gaps 25.

QY 9 KEEHEPEPTKAEVQGLPTWLOGVLLRNGCMHTIGDTKYNNHWFQGLALHSPTKN 64

DB 14 FSNVAVEFLPTFLFSSQIFALIKSIVRFGKRSQSLKGVHWFQGLALHAFHGE 73

QY 65 GEVYPSKYLPSDTYN-NIEANFVVSFEGTMAYPQPKNIFAKAFSYLSHT 123

DB 74 GAAELFYV-LGSAVALLFAGSYTCHQDMV 11 GCVVAVWFL 11 GAVIFVAT 129

QY 124 CLINIMKIGLYVAISHNFIKIDPQLLELAKVIV-SKYVAVNLATSHPHYS AGNI 181

DB 130 --TSVMALPDLWTLWAGHPYALDLENLATICTFPGAGPQPSAPHPIDAVTGEI 187

QY 182 LNMCTISVKGPTKYVVFIFTSVVFKEKSTFHLFVVSFISPSSTLQPSYVIF 241

DB 189 FSIQ-----VDFNLNLYPLPQSNLKHRL-----KLSPTPTQSPMA 229

QY 242 ENYIVFTEPEKELDKALATAYIKSNWASLSFKHEKALWHPVLPKAKASVSKFYD 401

DB 230 GPYLVLPLPTITNQESL-----LGNVAYALQAWAGQTEFVVFIPQSL 285

QY 402 ALVLYHHINAYE-FICHVWFVIVAYVYVINSLYLWETLKLKLEQEVSNF 460

DB 286 SWFQWYHNGCELDGCVNLGFFVFNFA-----QINLEYLRVIT 426

QY 216 KHELVVCSIPSKSLQNSYHSHFCITENYIVFIDQFPL---DIVKLATAYINGVWASC 272
 DB 255 MHDVPPTIS-----EPIMHDFAITETYAIPMLDMLHHPKEMVK-----EKKM1 300
 QY 273 LSFHKEDKTWFHVDRAKTKKEVSTK-FYTDALVLYHHINAYEEDGHVFDIVAYRNSLY 331
 DB 301 YSFDPYTKARQGVLPYAKUDELMPWFELNCFIFINANAEDEEVL-ITCRLENPDL 359
 QY 332 DMFYLLKLDKDFEVNNKLTSIPTCKRNVVPLQYDKDAEVSNLVKLPTSATAVEKDGSI 391
 DB 360 DMVSGKVKLEKLENGNELYEM-----RF-----NMKTGSA 389
 QY 392 YCQPELICEGLPRVAVNYGKKYKYVYATEVQWSPVPTKIAKLVQV-----440
 DB 390 -SQKLSASAVDFPRNRYTGKKQRYVYGTILUSTAKVGTIKPDLHAEAEKGMLEV 448
 QY 441 ----KEVLHMGEDHCWSEPIFVSPDAREDEGVLTICVVSSEPNKAPFLILLDAKTEK 496
 DB 449 GGNIKGIYDGEGR-YGSEALYVPEFA-BEIDGCLIFFVHDENIGKS-CVTIVIDAKIMS 505
 QY 497 ELGRATVNV--FMHLDLHGMFIPNDLCAET 525
 DB 506 AEPVAVVELPHRVPGYGFHALFVTEQLEQQT 536

RESULT 8

S76206

hypothetical protein - Synchocystis sp. (strain PCC 6804)

C:Species: Synchocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision: 25-Apr-1997 #text_change: 08-Oct-1999

C:Accession: S76206

P:Requester: T.; Sato, S.; Katsui, H.; Tanaka, A.; Asamizu, F.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res 3, 109-116, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis

S.

A:Reference number: S7422; M019904201; PM19905241

A:Accession: S76206

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-480 <KAN>

A:Cross-references: EMBL:D90914; GB:AH001349; NID:q1654477; PID:HAAL18465.1; PID:d101

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match

Best Local Similarity 23.4%; Score 280; DB 2; Length 480;

Matches 113; Conservative 90; Mismatches 210; Indels 70; Gaps 19;

QY 5 FNRKKEPEPTKAEVQGLPTWLOGVLLRNGCMHTIGDTKYNNHWFQGLALHSPTKN 64

DB 14 FSNVAVEFLPTFLFSSQIFALIKSIVRFGKRSQSLKGVHWFQGLALHAFHGE 73

QY 65 GEVYPSKYLPSDTYN-NIEANFVVSFEGTMAYPQPKNIFAKAFSYLSHT 123

DB 74 GAAELFYV-LGSAVALLFAGSYTCHQDMV 11 GCVVAVWFL 11 GAVIFVAT 129

QY 124 CLINIMKIGLYVAISHNFIKIDPQLLELAKVIV-SKYVAVNLATSHPHYS AGNI 181

DB 130 --TSVMALPDLWTLWAGHPYALDLENLATICTFPGAGPQPSAPHPIDAVTGEI 187

QY 182 LNMCTISVKGPTKYVVFIFTSVVFKEKSTFHLFVVSFISPSSTLQPSYVIF 241

DB 189 FSIQ-----VDFNLNLYPLPQSNLKHRL-----KLSPTPTQSPMA 229

QY 242 ENYIVFTEPEKELDKALATAYIKSNWASLSFKHEKALWHPVLPKAKASVSKFYD 401

DB 230 GPYLVLPLPTITNQESL-----LGNVAYALQAWAGQTEFVVFIPQSL 285

QY 402 ALVLYHHINAYE-FICHVWFVIVAYVYVINSLYLWETLKLKLEQEVSNF 460

DB 286 SWFQWYHNGCELDGCVNLGFFVFNFA-----QINLEYLRVIT 426

A:Reference number: 221500

A:Accession: T34293

A>Status: preliminary; translated from Gb/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1760 nt

A:Cross-references: EMBL:U53341; PIDN:AAC6q106.1; GSPDR:GN00028; CFSP:F49E10.2

A:Experimental source: strain Bristol N2; clone F49E10

C:Genetics:

A:Gene: GSP:F49E10.2

A:Map position: X

A:Introns: 51/1, 32/3, 176/3, 235/3, 332/2, 514/1, 543/2, 503/3, 677/1; 732/3

Query Match

Best Local Similarity 9.5%; Score 269; DB 2; Length 790;

Matches 136; Conservative 91; Mismatches 221; Indels 154; Gaps 28;

QY 5 FNRKKEHPEPIKAEVQGL-----PTWLGVL-----RNGPMHTIGDTKYNHWD 52

DB 20 FNTHLGHPAAMDGKYRELYCFSKNIPKWLGDYFLCQLSASYGNSSAPEGE-KLNHMD 78

QY 53 GLALLHSFTFKNGEVYYSKYLRS-----DTYCNLI-----EANRIVSEFGT 95

DB 79 AIGAVGSHVNSGQVFSQAQYYPAPYK IWEFTDRNKSASVPWAGWSYDNLTAWSRWEQ 138

QY 96 M-AYDPCKNIFAKAFSYLSHTIPEPTDNCIJNIMKTGDDYATSETNFI--KKIDPQTL 152

DB 139 VPANPSAK-----FHPN--LDFWKVGNRIVAGTEAPYVWGVYEDVRTL 180

QY 153 ETLKDVYSK-----YAVNLTASHPHYDSAGNI-----LN-----MGTSI 188

DB 181 QKFLFPFKENDFISTPRHTMIPISMNI-HERNDAGTIGWSFSAMNFEEORFGQIFT 239

QY 189 VDKRTKVLFPKIPSSVPEKEKKKCFKHLVVCSIPSRSLPSVYHSGFITENYIV-- 246

DB 240 VDTNCRVRVGLIDYGVWD---INAGSSNEY---IGDKTL-PCYIHSITSTENFIILP 292

QY 247 -----FIFOPFKLIDIVKATAYIRGVNWSCLSPHKEDKTHFVDRKTKKEVST 296

DB 293 ITSLLINCKFKKEPLN---NVSRAIQKGLMG--MDFYDMVPMRFLFNKTKLEFTTS 346

QY 297 K--EYTDALVLYHINAYE EDGHVVFIDIVAYDNSLYDMFYLLKLDKDFEYNNKLTSP 353

DB 347 KPLEVFSMPVTHOLNAFEADDGNFVADVMVYDSDHP---YVKYPTDFLTQLYPSTA 402

QY 354 TCKREVPVLPQDKDAEVSGLNLPLTSATAVEKKGSIYQCPILCEGIELPRVNYDNG 413

DB 403 RVLRFET--LDSKKQRMVNYLVLPQETIAA-----DFPQFNHYEQ 440

QY 414 KKYKVVYATEVQWSIVPIKIAKLVN-----QTKEVLHWGEDHCWPSRPIFVP 460

DB 441 FPYLWGYLVQHPFAS--GNSITFINVPPAGNPNLEFPAPETTLVLIH-----FWPVL 490

QY 441 SPDAFFHNPVVLTCVVVSFNPFAPIILIDAKTFKELQAPATVNVEMHLDHGMFIQND 520

DB 491 KPSKKKEDGKGVLIWPGVLTAFNK7-VLIVTAFKMWELGPAVVPISIPSFHNPFFSKK 549

QY 521 LG 522

DB 550 LG 551

RESULT 12

AC2417

hypothetical protein all4895 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AC2417

R:Kuroki, T.; Nakamura, Y.; Wolk, C. P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iziguchi,

Nakazaki, N.; Shimojo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2417

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1472 nt

A:Cross-references: Gb:BA000171; PIR:BA00594.1; PID:q17134034; GSPDR:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4895

Query Match

Best Local Similarity 9.4%; Score 264.5; DB 2; Length 472;

Matches 109; Conservative 72; Mismatches 192; Indels 107; Gaps 17;

QY 20 VQSGIPWIGLGVIPNIPGPMHIIGDTKYNHWFETALIHSPFKNKFVYVPSKYLIPSDLY 79

DB 32 TSGNIPGLIGSTLYPNPAPIPWGGMLAGHWFETGAILAVNFTAGGVIAITKYVQIVQY 91

QY 80 NCNIFANPIVVSPEPTMAYPDDCKNIPAKAFSVISHTDEPTFNCITNIMKTGDDYAT 139

DB 92 QVENTADKELYGVYG-MTAPGVVWQWRKPIKNAANTSVLALFDKLLALWEGRPY 146

QY 140 ETNFIPIPDQTLFTLTKVPSKYVAVNLTATSPHY---NSA--NTINMATSIVWGRUKY 196

DB 147 -----ALDLOTLETTGWDNLGG--LTNGLTYSARYKDDPKETFPNPGISLNAK 195

QY 197 VLFKIPSSVPEKEKKKCFKHLVVCSIPSRSLPSVYHSGFITENYIVFI 248

DB 196 -----LNVYKSDSTGKTILOKSAYSLSLGLPLIHDFVLACKYLIFP 244

QY 249 EGPFKLGI--VKIATAYIRGVNWSCLSPHKEDKTHFVDRKTKKEVSKRYTIDALVLY 406

DB 235 ISPVRLANLPLVLLGTS-----NYSDSMQWRPELQTLIVIDRESLSLVS RGETELWYQW 288

QY 307 HHINAY--EECHVVVFQIVAYPENSLSYDMFYLLKLDKDFEYNNKLTSTPTCKRFVVP 465

DB 289 HFANGYVDASGAVLVDIARY-----KDFQTN-----QVL 417

QY 366 KDAEVSNI--VKLTTSATAVK--EFGSTVQCPPIFGFIFIPVNYIVNKKYKVVY--- 420

DB 318 KEVATGOTHTPAUTSTLSPVNLHPQSKVAALFGLDHPKCFEHPVPRQNVQASRYTMSG 477

QY 421 ---ALFVQNSVPKPIAKLVNLIKEVLH-WCHDHCWPSRPIFVPSIDAEDEKGVLLICV 476

DB 378 FROGTDISQELLNT--TACFDKRTOTFTIAYGNCYSPSEPIIA-----DNWVLIVV 427

RESULT 13

T51936

probable 9 cis epoxycarotenoid dioxygenase [imported] - potato

C:Species: Solanum tuberosum (potato)

C:Date: 20-Nov-2000 #sequence_revision 20-Nov-2000 #text_change 20-Nov-2000

C:Accession: T51936

R:Burridge, A.; Taylor, T.B.; Thompson, A.

submitted to the EMBL Data Library, March 2000

A:Description: potato putative 9 cis epoxycarotenoid dioxygenase 1 cDNA.

A:Reference number: 225874

A:Accession: T51936

A>Status: preliminary; translated from Gb/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-604 nt

A:Cross references: EMBL:AJ276244; FIDN:CA876920.1

C:Genetics:

A:Gene: need1

Query Match

Best Local Similarity 9.3%; Score 264; DB 2; Length 604;

Matches 128; Conservative 94; Mismatches 219; Indels 154; Gaps 27;

QY 1 METIFPNKKEEHEPEPIKAE-----VQGLPTWLGQVLLKNGHG 48

DB 104 VEGALTKHELFEPKPTADEFVGLSGNFAPVFNPPVQSLIVTKKLTICVGVVYVNRKAN 164

QY 39 MHTIGDTKYNHWFETGALLHSFTFKNGEVYYSKYLRSPTLYNCTNFANKIVVSEFTIMAY 98

[illegible]

```

RESULT 15
JC4324
Lignostilbene alphabeta-dioxygenase (EC 1.13.11.4) [I] isozyme beta chain - Pst
C:Species: pseudomonas paucimobilis
C>Date: 04-Jan-1996 #sequence_revision 08-Feb-1996 #next_change 17-Mar-1996
C:Accession: JC4324
F:Famada, S.; Fukui, Y
B:Biosci. Biotechnol. Biochem. 59, 1866-1868, 1995
A>Title: Cloning of a lignostilbenedioxygenase, beta-dioxygenase isozyme gene from P
A.Reference number J-4324, MIM: 646893, FGI: 8534977
A:Accession: JC4324
A:Molecular type: DNA
A.Features: 1 485 FAM>
A:Experimental source: TMV1099
A>Note: The translation initiation codon ATG is not given in this paper
C:Comment: This enzyme is significant in the metabolism of dimeric lignin compounds
C:Genetics:
A:Gene: IsdB
C:Keywords: oxidoreductase

Query Match      8 78; Score 245.5; DB 2; Length 480;
Best Local Similarity 22.4%, Pred. No. 9.4e-10;
Matches 111, Conservative 76, Mismatches 211, Indels 131; Gaps 22;

QY 19 EVCEALPLWLDGVLLPNPCHHIDFDITNHEWGLATIHSHFFENRFEVYSKYLRSDT 78
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
26 EWESEIPACILDEFPYVHPVAQLPEPPPEEPEDPDNDHGMVSLEKPKHGKLTDERGVADIK 85
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Q7	79	YNCNEANETVVSFQPMATVTCNNLFAKATYLSHTIFETPRLI	NI	128
Q8	86	WVFFPKAG--SLFG--AYENP		122
Q7	129	MKIGDDYVAISENIETKIDPQITLILDKLSAVAVNLAISHHYEVS--AGNIIMGTS	187	
Qb	123	WVAACKLVAMKEDESPGLIMFLEETETVTFNFCCKLRGTFSAHAKIDPVTNCFRCGYA	182	
Q7	188	IVDKGTKYVLKPIPSSVPEKEKKSCFKHLEVVSIPSPLI-----QPSYV---HSPG	249	
Qb	193	AAGLLIIGCSVELD	FAKRLPFLHLLQVAYGGMHLYG	250

240	ITENYIVFIEOPKELDIVKLATAYITGVNNAWASCLSEHKEDKTFWIFVDR--- <td>296</td>	296
Db		
221	LTETAYATP-----HIVFGSPMKRIFAGLTHFEFETTLFVNLGVVPGCVITNKDFEW	273
QY		
247	KFYDALVLYHHINAYREDGHVDFIVAYHGNLSLYDMFYLLKLDKDFEVNNKLTSIPTECK	356
Db		
274	EKAAPKTFASHVNNAFEEGSKIHFDTHAENNAF---PFPDTHGAPFD-----PVAA	323
QY		
457	KFVVPDQVQKTAFAVSNIVKLPTSATVAKFKVSSVQBPTEICST-ELPKNVNYVNNCK	415
Db		
324	R---PYLRHWITVGLGSN-----SEDFAEVKGQITSWIDEFPFVADARYVQGP	365
QY		
416	YKYVYA-----FVQVSSPVPIKILNVVILKEVLHGEDH-----QWP-----SEP	456
Db		
466	YHHCWGLVMQDEMEEFARGASGFKN-----RIGHW---DHATGKEDSWMCCPOSIIQEP	419
QY		
457	IFVTSIDAREHDEGVLTVCVVSEPNKAPFLLILDAKTEKE---LGRATVNVNEMHIDJHC	513
Db		
420	CFVDFPMALNSAFRGYCIITALVDNIITNVSDI---VVDLNLNKDQICGPAKLPIRLRSGLHG	477
Db		

Search completed: July 15, 2003, 09:30:30
Job time : 46 secs

Result No.	Query			ID	Description
	Score	Match	Length		
1	121	4.3	1861	1	APU_THETU
2	112	4.0	749	1	PEX_HUMAN
3	110.5	3.9	1228	1	YHV5_YEAST
4	110.5	3.9	3829	1	SACS_HUMAN
5	108.5	3.8	2136	1	YFP2_MAPPA
6	108.5	3.8	3224	1	RFP2_HUMAN
7	107	3.8	541	1	YHFX_ECGF7
8	107	3.8	798	1	NP92_F0PVP
9	107	3.8	841	1	KFX4_DKQWE
10	107	3.8	1211	1	DP3A_HELCPY
11	106.5	3.8	1742	1	M5C_HUMAN
12	106	3.8	994	1	DP02_KLGLA
13	105.5	3.7	749	1	PEX_MOUSE
14	104	3.7	1581	1	AP01_PNECA
15	103.5	3.7	674	1	DTF1_MSEPV
16	103	3.6	587	1	RTFS_PLARA
17	103	3.6	804	1	SCY1_YEAST
18	103	3.6	900	1	GRAH_BAUSD
19	103	3.6	1042	1	SVL_FERRR
20	103	3.6	1539	1	SMCY_HUMAN
21	102	3.6	402	1	P01R_VICBR
22	101.5	3.6	1447	1	B2D4_YEAST
23	101.5	3.6	3112	1	LMA_ZEOMF
24	101	3.6	967	1	T2S7_ECOLI
25	101	3.6	1865	1	DPVA_STAAU
26	100.5	3.6	796	1	TLR6_HUMAN
27	100	3.5	883	1	YGLB_SILPTT
28	100	3.5	969	1	VGAU_SHIMP
29	100	3.5	999	1	DSC3_HUMAN
30	100	3.5	1211	1	DP3A_HELCPJ
31	99.5	3.5	893	1	PRIM_FITPP
32	99.5	3.5	1976	1	MYHA_BOVIN
33	99.5	3.5	1976	1	MYHA_HUMAN

DR pfam: PF00128; alpha-amylase; 1.
 DR pfam: PF00395; SLH; 3.
 DP pfam: PF02806; alpha-amylase_C; 1.
 DP pfam: PF02903; alpha-amylase_N; 1.
 DR pfam: PF02922; isomylase_N; 1.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PS01072; SLH_DOMAIN; 3.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
 KW Multifunctional enzyme; Glycoprotein.
 FT SIGNAL 1 35
 FT CHAIN 36 1861
 FT DOMAIN 428 1018
 FT DOMAIN 1157 1248
 FT ACT_SITE 628 628
 FT ACT_SITE 657 657
 FT ACT_SITE 734 734
 FT ACT_SITE 734 734
 FT DOMAIN 1681 1739
 FT DOMAIN 1740 1803
 FT DOMAIN 1804 1861
 FT CONFLICT 1734 1734 D -> E (IN REF. 1; AAB00841).
 SQ SEQUENCE 1861 AA; 206104 MW; 30625070E4548574 CRC64.

Query Match 4.38; Score 121; DB 1; Length 1861;
 Best Local Similarity 20.28; Pred. No. 1.3;
 Matches 117; Conservative 91; Mismatches 224; Indels 146; Gaps 30.

QY 23 QLPTWLGVLRLNGPG-----MHTIGDKYNHW-----FDGLAL-LHSFTFKNGEV 67
 DB 931 CAPS:---NVVTSNGKVLDSWLSGATYNYKSSVEGLYKLSNVTTTFEDANV 987
 QY 68 YRSKYLRSPTYNCINIANRIVSEFG---TWAYDPCKNIFAKAFSYLSHTIPEFTDN 123
 DB 988 TNLKLYVYA-----ISAIDELNCSISNDAYAP-----AYPIGWGN-LTQVSDN 1033
 QY 124 CLINIMTKGDYATSETNFIKIDPOTLETLDKVDYKYVAVNLTATSPHVDSA-GNTL 182
 DB 1034 HIGVDKPTDIYAEVWADGLTNGTQCPNMIAQLGY-KYVSTV-----YDSVVGSVY 1086
 QY 183 NMGTSIVDKGRTKYVLFKIPSSVPEKEKKCKHLEVCISPSRLLQPSYHSHFG---- 239
 DB 1087 NSVYGVDDSGFT-WVNAQYVGDIGNDQYKASF-----TPDK-IGQWEVLMRFSNQ 1136
 QY 240 ----ITENYIVFIEQPEKLDIVKLATAYIRGVNWSCLSFHKEDKTFHFVDKTKKEVS 295
 DB 1117 GDMWITTSLSFYVVPD-DLIKPTAPYLNQPTES-----SRVSIW----- 1178
 QY 296 TKFYTDALVLYHHINAYEEDGHVVOIVAYRDNLSYD-----MFYLLKKDKDFEV 345
 DB 1174 NPSTDNVGVIVDY-EIYPSDGG-TFNKIATVSNEVINYDTSVINGVTYNYKVVAVDLSF 1235
 QY 346 NKKLTSIPTCKRFVVPLO-----YDKDAEVSGLV-KLPT-----SATAVKEKDG 390
 DB 1236 NKTESNVVTKPDVVPVPIKVFNTVDPDTPCA---VNLASTFPNATWDSAGQMTKIDNN 1292
 QY 391 IYCQPEILCEGIELPRVNYDNGKYYKYVATEVQSPVP-----TKIAKLNVQ 439
 DB 1293 TYSITLIDETQI-----EKKYARGS---WDKVEKDEYNEFASNRKVTIVNOG 1339
 QY 440 TKRY:HWFFDHCWSEPTFVPSDPAPEE:EGVYVLTCCVVSSEPNKAPFLILDKATKE-- 497
 DB 1340 NNEMTINDTVYWRWDIPFIYSPSSNMTVDSNISTMEVKGNTYKGAKVTI-NGDSFVQDK 1398
 QY 498 -----LGRATVNVEMHLDLHGMPIDNDLGAETE 526
 DB 1399 NGVTKVDSLVNGYGVNKKIKIHVEPNDSGVYNGNDGGRITE 1436

RESULT 2
 PEX_HUMAN
 ID PEX_HUMAN STANDARD: PRT: 749 AA.
 AC P78562; 000678; 09827; Q13646; Q93032;
 DT 01-NOV-1997 (Ref. 35, Created)
 DT 01-NOV-1997 (Ref. 35, Last sequence update)

DI 15-JUN-2002 (Ref. 41, Last annotation update)
 DE Phosphate regulating neutral endopeptidase (EC 3.4.24.-)
 DE (Metalloendopeptidase homolog PEX) (X-linked hypophosphatemia protein)
 DE (HYP) (Vitamin D-resistant hypophosphatemic rickets protein).
 GN PEX OR PEX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS HYP ARG-85; LEU-514; ARG-579 AND
 RP PRO-651.
 RX MEDLINE=97343325; PubMed=9199930;
 RA Francis F., Strom T.M., Hennig S., Boeddrich A., Lorenz H.,
 RA Brandau O., Mohlke K.L., Cagnoli M., Steffens C., Klaus S.,
 RA Borzay K., Mohl T., Oudet C.L., Econs M.J., Powe P.S.N., Reinhardt R.,
 RA Meitinger T., Lebrach H.;
 RT "Genomic organization of the human PEX gene mutated in X-linked
 RT dominant hypophosphatemic rickets.";
 RL Genome Res. 7:573-585(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97232252; PubMed=9077527;
 RA Beck L., Soumounou Y., Martel J., Krishnamurthy G., Gauthier C.,
 RA Goodyer C.G., Tenenhouse H.S.;
 RT "PEX tissue distribution and evidence for a deletion in the 3'
 RT region of the Pex gene in X-linked hypophosphatemic mice.";
 RL J. Clin. Invest. 99:1200-1209(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone;
 RX MEDLINE=97343443; PubMed=9199999;
 RA Guo R., Quarles L.D.;
 RT "Cloning and sequencing of human PEX from a bone cDNA library:
 RT evidence for its developmental stage-specific regulation in
 RT osteoblasts.";
 RL J. Bone Miner. Res. 12:1009-1017(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Lipman M.L., Panda D., Henderson J.E., Shen Y., Goltzman D.,
 RA Karaplis A.C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97224400; PubMed=9070861;
 RA Griff M., Mumm S., Waeltz P., Mazzarella R., Whyte M.P.,
 RA Thakker R.V., Schlessinger D.;
 RT "Expression and cloning of the human X linked hypophosphatemia gene
 RT cDNA.";
 RL Biochem. Biophys. Res. Commun. 231:635-639(1997).
 RN [6]
 RP SEQUENCE OF 4-641 FROM N.A.
 RX MEDLINE=96024647; PubMed=7550339;
 RA Francis F., Hennig S., Korn B., Reinhardt R., de Jong P., Poustka A.,
 RA Lebrach H., Powe P.S.N., Goulding J.N., Summerfield T., Mountford P.,
 RA Read A.P., Popowska E., Pronicka E., Davies K.E., Orford J.L.H.,
 RA Econs M.J., Nesbitt T., Brezner M.K., Oudet C.L., Pannetier S.,
 RA Hanauer A., Strom T.M., Meindl A., Lorenz H., Cagnoli M.,
 RA Mohlke K.L., Murken J., Meitinger T.;
 RT "A gene (PEX) with homologies to endopeptidases is mutated in
 RT patients with X-linked hypophosphatemic rickets. The HYP
 RT Consortium.";
 RL Nat. Genet. 11:130-136(1995).
 RN [7]
 RP SEQUENCE OF 1-116 FROM N.A.
 RA Waterston R.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP VARIANTS HYP TYR-85; CYS-166; SEP-252; ILE-254 AND VAL-579.
 RX MEDLINE=97260404; PubMed=9106524;
 RA Holm I.A., Huang X., Kunkel L.M.,
 RT "Mutational analysis of the PEX gene in patients with X-linked
 RT hypophosphatemic rickets.";

DB 414 MMEELIVEGVWAFIDMLEKENE-----WMDAGTKRKAKEKAPAVLAKGVYPERIMNDTHVN 489
 QY 411 AYEHGHHVVDIVAYRONSLY-----DMFYLLKLDKDFEVNNKLTISPTCKRFVPL 362
 DB 490 --EDLKAKESEADYFGNVLQTRKYLAQSDFFWLR-----AVPKTEWFTNPT 535
 QY 464 QYDKDAKGVSNLVKLPTSATAVKKGDSIYQPEILCEJIEUPRV----- 407
 DB 516 TVNAYFASINQIRFPAG-----ELUKPFW-----GTEPRSLUSYGAIGVINGHEPT 583
 QY 408 --NYDNGKKY-----KYVYATEV-----QWSPVPTKIATLNVOITKEVLHWG 447
 DB 584 HGFUNNGKRYDKNGLDPMWSESEKFKKTKCMINQISNYWYKAGLVNKGKRTL--G 641
 QY 448 ED 449
 DB 642 EN 643

RESULT 3

ID YHV5_YEAST STANDARD: PPT: 1228 AA.
 AC P38851;
 DT 01-FEB-1995 (Rel. 31, last sequence update)
 DT 15-JUL-1998 (Rel. 36, last annotation update)
 DE Hypothetical 143.6 kDa protein in SP016-PEC104 intergenic region.
 GN YHR155W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.
 OC Saccharomycotales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN S298C / AG072;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnson M., Andrews S., Brinkman K., Cooper J., Bing H., Dover J.,
 Du Z., Favetto A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 Nhan M., Pitkin L., Piles I., St Peter H., Trevaskis E., Vaughan K.,
 Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 Vaudin M., Vaudin M., Vaudin M., Vaudin M., Vaudin M., Vaudin M.,
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII."
 RT Science 265:2077-2082(1994).
 RL Science 265:2077-2082(1994).
 CC -1- SIMILARITY: STRONG. TO YEAST STP3 SAME. TO S POMRE SPAN19AR 02
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb.sib.ch/announc/>
 CC or send an email to license@isb.sib.ch).

CC FMRL: 010347; AAF68977.1;
 DR PIR: S46754; S46754.
 DR SGD: S0001198; YHR155W.
 DR InterPro: IPR001849; PH.
 DR Pfam: PF00169; PH; 1.
 DR SMART: SM00233; PH; 1.
 DR PROSITE: PS50004; PH_DOMAIN; 1.
 KW Hypothetical protein.
 FI DOMAIN 408 421 PH.
 SO SEQUENCE 1228 AA; 143583 MW; C8872FAEB270A4B6 CRC64;

Query Match

Best local similarity 84.8%; Score 110.5; DB 1; Length 1228;

Matches 88; Conservative 78; Mismatches 150; Indels 151; Gaps 23;

QY 58 HSFTFKNKVEYVRSKYLRSIDLYNENLKANKIVVSEKAGIMAYDPCNKFAPKAFSYLSHII 117

DB 571 YPSLKVDTQPSISFF-SVNHDFLQVPLVILPSSVWTPNNKGFASMAVITNIIIV 629
 QY 118 PEETDNCINIMKTGGDYATSETNFIPIKIDPOTLETIKVIYSKYVAVNIATSHDRVHS 177
 DB 630 V-----YLNISGFYSYLRKID--LIDIDHSIEYDK - SEKHVSS 662
 QY 178 A-----UNILNMGISIVK-----GRKYVVLKIPSSVPEKKEK-KSTFAHLE- 219
 DB 663 RMLHMGGHICLRFNMVSVEIFDRZAVASKELFLENAMHIFKRFKVLIFPEHJLFFIPN 722
 QY 220 ---VVCISPKS-----LQPSVYHSKSTENIVFLECFE-----ELHIVKLAIA 262
 DB 723 EKLIKDNSEHSEHYSKDYDYLKSYDHHFNTN-----ETPMELMSKKLRLEHAW- 776
 QY 263 YIPGVNWAACISFHKDKTWFH--FYDPTKKEVSTKFTYTDALVL-----YHH 408
 DB 777 YFQD-----NEKVGSKTLFHLFGDK-----SQVFPSSLCKKGSNNNNNSYWER 822
 QY 309 INAYEEDGHVVEDI-----VAYRDNLSYDMFLAKLKDQFE--VNNKLTSTPTCKRFV 359
 DB 823 IRPAKEDASQFELCPKQLQFNSTNFIKDLLWLKNDNFKLVLQVPTKIKO----- 877
 QY 360 VPIQYDKDAEVSGLNVLKLPSTATAVKKGDSIYQPEIL-----CRGIEIPR - 406
 DB 878 --GYFEFVEKGPILVKFP-----LCHPLLLKVKRFLIAAIISSQESLAKGTD 921
 QY 407 --VNYDNGKKYKYVYATEVQWSPVPTKIATLNVOITKEVLHWGCEDH- 451
 DB 522 LATIYDFN YVESID FIKETKIMLIFERH:NMALFYC 959

RESULT 4

SACS_HUMAN STANDARD: PPT: 3829 AA.
 ID Q9NZJ4; Q94835;
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Sacsin.
 GN SACS OR KIAA0730.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND VARIANT ALA-2619.
 RX MEDLINE=20120709; PubMed=10655055;
 RA Endert J.C., Herube P., Merleier J., Dore C., Lepaque P., Gu B.,
 Richard J.-P., Mathieu J., Molancon S.B., Schalling M., Lander E.S.,
 Morgan K., Hudson T.J., Richter A.;
 PI "AFSACS, a spastic ataxia common in northwestern Quebec, is caused by
 PT mutations in a new gene encoding an 11.5 kb ORF."
 PL Nat. Genet. 24:120-125(2000).
 EN [2]
 RP SEQUENCE OF 2825-3829 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.I., Suyama M., Kikuno R., Miyajima N.,
 Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286(1998).
 CC -1- FUNCTION: MAY FUNCTION IN CHAPERONE-MEDIATED PROTEIN FOLDING.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE CENTRAL NERVOUS
 CC SYSTEM. ALSO FOUND IN SKELETAL MUSCLE AND AT LOW LEVELS IN
 CC PANCREAS.
 CC -1- DISEASE: DEFECTS IN SACS ARE THE CAUSE OF AUTOSOMAL RECESSIVE
 CC SPASTIC ATAXIA OF CHARLEVOIX-SAGUENAY (AFSACS OF SACS). AFSACS IS
 CC AN EARLY ONSET NEURODEGENERATIVE DISEASE WITH HIGH PREVALENCE IN
 CC THE CHARLEVOIX-SAGUENAY-LAC-SAINT-JEAN REGION OF QUEBEC. IT IS
 CC CHARACTERIZED BY AHSNI SENSORY-NEURVE CONDUCTION, REDUCED MOTOR

RA Grotbeck E.J., Davis N.W., Lim A., Dimelanta R.T., Potamoumis K.,
RA Apodaca J., Anantharaman T.S., Jin T., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
PL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N A
RC STRAIN-0157:H7 / PMID 0509952;
PX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
FT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
PL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: INVOLVED IN ADHERENCE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (Probable).
CC -!- SIMILARITY: BELONGS TO THE UPF0141 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U50906; AA06312.1; ALT_INIT.
DR EMBL; AE005546; AG58308.1; ALT_INIT.
DR EMBL; AP002564; BAB37476.1; ALT_INIT.
DR InterPro; IPR003371; DUF146.
DR Pfam; PF02418; DUF146; 1.
KW Transmembrane; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 25
FT CHAIN 26 541
FT TRANSMEM 61 81
FT TRANSMEM 111 131
FT TRANSMEM 142 162
FT TRANSMEM 265 285
FT SEQUENCE 541 AA; 60639 MW; 89C140B666FA0894 CPO64;
SQ
Query Match 3.8%; Score 107; DB 1; Length 541;
Best Local Similarity 20.6%; Pred. No 2 5;
Matches 80; Conservative 58; Mismatches 130; Indels 120; Gaps 19;
QY 64 NGEVYRSKYLRSITYNCNEANRIVVSEFGTMYDPCKNIFAKA-----PSYLSHTI 117
DB 152 SGLSFCACQFAYKDAKKNAPSPYILASRFATYT-PFENLNYPALAAKDHORLLSTANTV 210
QY 118 PEFTDNCNLINIMKTGDYYA-----TSETNFKRIDPOTLETLDK-----V 158
DB 211 PYFQ-----LSVRDGTIDTYLVIVGESVRVDMNSLYGYTRSTPTQ-VEAQRKQIKLENQAI 265
QY 159 DYKYVAVNL-----ATSHPHYSAGNILNM-----GTSIVD 190
DB 266 SGAPYITALSVPLSLTADSVLSHDITHNYPDNIINMANQAGFOTFWLSSQSAFRQNGTAVTS 325
QY 191 -----KGTQYVLEKIPSSVPEK-EKKKSCFKHL-----EWCSCI--PSRSLQOP 232
DB 326 IAMPAMETVYVPGFDELLPHLSALQOONTQKKLLVLHLNGSHPEACSAFYQSSAAYFQP 385
QY 233 S-----YHHSFGITENYI-----VFTEQPFKLDIVKLTATAYIRGVNNAAS 271
DB 386 QDDQADACYNDSIHYTDSLLQGVPELLKDRASVMYFADHCLPPTTK-KNVYFHCGFAS 444
QY 272 CLSFHKEDKTFW-----HFVDRKTKKEVSTKFTYDVALVYHHINAY-----EEDCHVVF 320
DB 445 QQAYHVPWFNIWSPVLGVDGVDRTENNIFSTANNYL-----INAMMGVTKEQPQTLRF 499
QY 321 DIVAYRNSL-----YDMFYLLKLDKDF 343

DB 500 VIVHYKGDLSVVDANHDVFDYVVLKPEF 527
RESULT 8
ID RP94_FOWPV STANDARD; PRT; 798 AA.
AC Q9J589;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PNA-polymerase-associated transcription specificity factor (PAP94).
GN FV141.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID:10261;
RN [1]
RP SEQUENCE FROM N A.
EX MEDLINE-20193820; PubMed-10725156.
RA Alfonso C.L., Tulman E.R., Fu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus";
PL J. Virol. 74:3815-3831(2000)
CC -!- FUNCTION: ASSOCIATES WITH THE PNA POLYMERASE, REQUIRES FOR THE
CC TRANSCRIPTION OF EARLY GENES. POSSIBLY MEDIATES THE BINDING OF
CC THE CORE PNA POLYMERASE TO VETF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF198100; AAF44485.1;
DR InterPro; IPR004974; Pox_Pap94.
DR Pfam; PF01294; Pox_Pap94; 1
KW Transcription; Late protein.
SQ SWISS-PROT 798 AA; 85320 MW; 2AEECF453537B03 CPO64;
Query Match 3.8%; Score 107; DB 1; Length 748;
Best Local Similarity 21.8%; Pred. No 4 3;
Matches 94; Conservative 54; Mismatches 171; Indels 102; Gaps 24;
QY 73 YLRSDT-----YNCNEANRIVVSEFGTMYDPCKNI-----FAKAF 110
DB 18 YIKDDTVKESYODFIEKNKELFICNLVNNMTDEIKLLYIIQNDIDLUKSLVAIF 77
QY 111 SYLSHTIPE--PTUNILINI--MKTDQYVAISEINPIPKIDPQTLFELFVIVSYAV 166
DB 78 SYIGYNEKNLHDUNSSLDGDRMIGDMNVMYDTFF-----SLDFTIRQKHVI 128
QY 167 NLATSHPHYSAGNILNMGTSIVDKGPKY--VLEKIPSSVPEK-----EKK 211
DB 129 -LVNDEGNDFNINYSFTTSLSYK-EDKYEQVNEIPFNMKELLSYVSKNLDQLRPSKK 186
QY 212 KSDPYHL--EVMSTFSPSLIATSYVHSFGTIERVYVFIPIQYFIIVYLAATAYIP 259
DB 187 YLDFAYLCLRNIGIKISKPK-YNVYIFVYVIDELTIPVIK-----DYLDVKVYILEETNK 241
QY 270 ASCLSTPIKEKR:WHEVFAKIKREVSTKPYTLALVLYHHINAYE HCHVVF 420
DB 242 AYPNPNFNQNKYPVFW-----GRVITPKFNPPILSYVFTISVYGIQIEMLTINP 296
QY 321 PIVAYPINSLYDMFYLLKFLKIKDFEVNKKLISLIPCKKPPVVPQYER-DAEVSNSLV 379
DB 297 E---PRKNPI-EYIYVSEL-KFWEEGGSVDVFP-CEHETAIIDAKKVSLEYEYENIN 348
QY 360 SATAVKREKISTVQPTIATGELFPVPPVNYFNWKKYKYVYATFVWSSVPIKIAKIN 439
DB 349 IAKYIYVEDGLAYCN---LC-GINIOELNLD-----ATDV-----TKISLINVT 388
QY 440 TKEVLHWGEDH 450

```

100 389 YNKSIFMSDEY 399
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```


entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

CC EMBL: AF272390; AAF78783.1;
CC HSSP: P08799; IYOM.
CC Genew: HGNC:7604; MYO5C
DR InterPro: IPR002710; DIL.
DR InterPro: IPR000048; IO_region.
DR InterPro: IPR001609; myosin_head
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IO; 6.
DR Pfam: PF01843; DIL; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1
DR ProDom: PD003376; DIL; 1.
DR SMART: SM00015; IO; 5.
DR SMART: SM00242; MYO5C; 1
DR PROSITE: PS00096; IO; 4.
KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Polymorphism.
FT DOMAIN 1 755 MYOSIN HEAD-LIKE.
FT MYOIN 756 779 IO 1.
FT MYOIN 780 806 IO 2.
FT MYOIN 807 829 IO 3.
FT MYOIN 830 854 IO 4.
FT MYOIN 855 884 IO 5.
FT MYOIN 884 1451 COILED COIL (POTENTIAL).
FT MYOIN 1574 1679 DILUTE.
FT VARIANT 522 522 P->L.
FT VARIANT 634 634 /FTID-VAR_010646.
FT VARIANT 634 634 L->S.
FT VARIANT 634 634 /FTID-VAR_010647
SQ SEQUENCE 1742 AA; 202733 MW; 25DD3082A7EB5AAB CRC64,
Query Match 3.88; Score 106.5; DB 1; Length 1742;
Best Local Similarity 20.08; Pred. No. 13;
Matches 90; Conservative 73; Mismatches 178; Indels 109; Gaps 21;
QY 60 FTFRNGEYVRS---KYLRSSTYN-CN--IFANFIVVSEFTMAYDPCKNIFAKAFSYLS 114
DB 562 FLENPTVTIMLVETILRASKFHCANFFQENPTTPSPESMTITVKSQVVKPKSKHFR 621
QY 115 HTIPEFTDNCILINIMTKGDDYYATSETNFTIRKIDPOTIETLQKVY---SKYVAVNI--- 168
DB 622 TTGSKPSSLYLLMET-----LNATPHYVRCIKPND---EKLPEFDSKRIVQOLRAC 673
QY 169 -----ATSHPHYDSAGNILLMOTSIIVDMGFTKYVLKIPSSVPEKEKKKCFKHLE 219
DB 674 GVLETFIRISQSYD-----SRWTYIEFYSRYGI-LMTKQELSFSDKK 714
QY 220 VVCSIPSRSLQPSYHSFGITENY--IVFIEQPFKLDIVK----LATAYIRGVNKA 270
DB 715 EVCKVVLRLIQDSNQYQFGTKITFRAGOVAYLEK-LRLDLKROSCVMVQKHMRG--WL 771
QY 271 SCLSHFKDKTWF-----HFVDRTKKE-----VSTKFTDALVLYHHI-----NAYEELG 316
DB 772 QRKFLRERRAALLIIOYFPGQOTVRKAITAVALKAWAAIILQHCRCGLYVRSLYQLIR 831
QY 317 HWVEDIVAYRDSNLYDMFYLLKILDKDFE-----NNKITSIPGKPFVPLQY- 364
DB 832 MATITMGAYSNGELAPRPYFKMIFEHKAVILQKYAPAWLARRFOSI---RRFVLNIQIT 888
QY 465 -----DKLAEVGSNLYKLPTSATAVEKDGSIYCOPEILCEGIELPRVNDYNG 413
DB 889 YRVORLQKLEQONKE-NHGLVKFLTSLAATPAQVTEKTKI FAFIEKAATHPNPEKNG 947
QY 414 KKYKYVYATRVQWSPVPTKIANKLVQTKV 443
DB 948 KKYR-----DAVEEKLAKIQKINSEL 968
RESULTS 12
DRO2_KIULIA

```

```

IU DRO2_KIULIA STANDARD; PRI: 994 AA.
AC P05468;
DI 01-NOV-1988 (Rel. 09, Created)
DI 01-NOV-1988 (Rel. 09, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (Plasmid pCKL 2 protein 2).
OS Kluyveromyces lactis (Yeast).
OS Plasmid pCKL-2.
CC Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
CC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.
CC NCBI_TaxID=28985;
CC [1]
CC SEQUENCE FROM N.A.
RN STAIN-CES 2359 / IPO 1267 / NPBL Y-1140;
RX MEDLINE=88289339; PubMed=3041369;
RA Tommasino S, Kiesel S, Galeotti C.L.;
FT Gene-ome organization of the killer plasmid pCKL2 from Kluyveromyces
FT lactis."
RL Nucleic Acids Res. 16:5863-5878(1988).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate
CC + {DNA}(N).
CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.
CC -1- MISCELLANEOUS: THE PRESENCE OF THE TWO LINEAR PLASMIDS, TERMED
CC pCKL-1 AND pCKL-2, IN STRAINS OF KLUYVEROMYCES LACTIS CONFERES THE
CC KILLER PHENOTYPE, I.E. PRODUCTION OF TOXIN AND RESISTANCE TO IT,
CC TO THE HOST CELL.
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE B FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is published through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its copyright is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: X07776; CAA30603.1; -.
CC PIR: S00940; S00940.
CC InterPro: IPR002064; DNA_pol_B.
CC InterPro: IPR004868; DNA_pol_B_2.
CC Pfam: PF03175; DNA_pol_B_2; 1.
CC SMART: SM00486; POLB; 1.
CC PROSITE: PS00116; DNA_POLYMERASE_B; 1.
CC Transferase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Plasmid.
CC SEQUENCE 994 AA; 117560 MW; 1278F30C2D07F1F5 CRC64;
Query Match 3.88; Score 106; DB 1; Length 994;
Best Local Similarity 18.68; Pred. No. 6.8;
Matches 101; Conservative 71; Mismatches 160; Indels 212; Gaps 27;
QY 65 GEVYYS-----KYLRSSTYN~NIFANFIVVSEFTMAYGDP~KNIFAKAFSYLSHTIPE 119
DB 389 GELYRTGKNCIKPEFVEELSYNVDI--ILYAWYGS-----GGDY QHVLVY 442
QY 120 FTNCL-----INIMTKGDDYYATSETNFTIRKIDPOTIETLQKVY---SKYVAVNI 170
DB 433 MKSKCIKDKYIIKNNMTYGLY---ENALYIKDPYILTSILKA-SKAFVINEG 487
QY 171 SHPHY-----ISANQDINM~TSIVDMGFTKYVLKIPSSVPEKEKKKCFKHLE 199
DB 488 EPDHSIHSWEDLNKILPNWVKIQKRMIEKYNKLNLYKNIETPENKNNIILENAI 547
QY 200 ---KIPSSVPEK---EKKSKCFEHLVW~SILSPSLQ-----PSYHSP 248
DB 548 EYCKVDVLAMEKVIKFKLLKANLIIVSVKFTLSQLSMKIMSKLIFKYYKLYVPILE 607
QY 239 -----GTTENVIVFIE---QPELIDIVKLA IAYI- 656
DB 608 EYSKIRNATVGGPVIKNGIYPIENVIVAVVSIYPSAMKILKHSYCKOKIVLYIIEFKHG 667
QY 267 VNNASCLSFH-KELKTWTFHFVDKTKVESTKPY-----THALVYHHI 409

```


Db 307 SPFNLFPHINTYEDHEFLIVDLCQWKREFVYNYLIANI FENWEVKNAKAKZFEVVF 366
QY 357 RVVPLQVQKAEVGSNVLKLP-ISAIAVKKEKUGSYQPEILCEG-----IELRVNYD- 410
Db 367 PYVPLNDIK-ADTGKKNLVTIPNTATAILSPETITWLEPEVIFSGPQAFEPQJINYCK 425
QY 411 YNGKKYKYVATEVOWSPVTEKAKLNVOIKVELVHGEDHCWSEPIFVPSDAREDEG 470
Db 426 YGKPYIYAYGLNLHF-VDRICKLNKTKETWVQEDSPSYSEPIFVSHDDALBEDG 484
QY 471 VLTGVVYVSEPNKAP-ELLIDAKTEKELGRATVNVEMHLDLHGMF 515
Db 485 VLSVWVSPGQKPAYLLILNAKDLSEVARAEVINPVTPLGLF 530

RESULT 4

US-09-385-259-3
; Sequence 3, Application US/09/385-259
; Patent No. 6201114

; GENERAL INFORMATION

; APPLICANT: Aguirre, Gustavo D.

; APPLICANT: Acland, Gregory M.

; APPLICANT: Ray, Kunal

; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS

; FILE REFERENCE: IN DOGS

; CURRENT APPLICATION NUMBER: US/09/385,259

; CURRENT FILING DATE: 1998-08-30

; EARLIER APPLICATION NUMBER: 60/103,219

; EARLIER FILING DATE: 1998-10-30

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 3

; LENGTH: 205

; TYPE: PRT

; ORGANISM: Canis familiaris

US-09-385-259-3

Query Match 13.2%, Score 273.5, DB 4, Length 205,
Best Local Similarity 49.3%, Pred. No. 1.1e-29,
Matches 72, Conservative 21, Mismatches 46, Indels 7, Gaps 2;

QY 2 ETIENRKEERPEPIKAEGGQLPTWLGQSVLLKNGPMHTIGDTKYNHWFDSLALHST 61
Db 17 FTV-----EELSSPLTAHVTCRIPRLWLTGSLRCGPGLEFVSGSEPYHLFDGQALLHKFD 71
QY 62 FKNFVYVPSKYPSSTYNYCNIEANPIVVSFFETMAYPPCKNIFAKASYLSHTIPEF 121
Db 72 FKSHVTVHPFFIPTTAYVPMATEFYFIVITEPTTAPPTCPNIFSPFSYFVGV--EVT 124

QY 122 DNCLINIMKTGDYYATSETFIRKI 147
Db 130 DNALVNVYPGEDIYACTEINFITLI 155

RESULT 5

US-09-645-370-3

; Sequence 3, Application US/09/645-370

; Patent No. 6428958

; GENERAL INFORMATION

; APPLICANT: Aguirre, Gustavo D.

; APPLICANT: Acland, Gregory M.

; APPLICANT: Ray, Kunal

; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS

; FILE REFERENCE: IN DOGS

; CURRENT APPLICATION NUMBER: US/09/645,370

; CURRENT FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/485,259

; PRIOR FILING DATE: 1998-08-30

; PRIOR APPLICATION NUMBER: 60/103,219

; PRIOR FILING DATE: 1998-10-30

; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-645-370-3

Query Match 13.2%, Score 373.5, DB 4, Length 205;

Best Local Similarity 49.3%, Pred. No. 1.1e-29,
Matches 72, Conservative 21, Mismatches 46, Indels 7, Gaps 2;

QY 2 ETIENRKEERPEPIKAEGGQLPTWLGQSVLLKNGPMHTIGDTKYNHWFDSLALHST 61
Db 17 FTV-----EELSSPLTAHVTCRIPRLWLTGSLRCGPGLEFVSGSEPYHLFDGQALLHKFD 71
QY 62 FKNFVYVPSKYPSSTYNYCNIEANPIVVSFFETMAYPPCKNIFAKASYLSHTIPEF 121
Db 72 FKSHVTVHPFFIPTTAYVPMATEFYFIVITEPTTAPPTCPNIFSPFSYFVGV--EVT 124
QY 122 DNCLINIMKTGDYYATSETFIRKI 147
Db 130 DNALVNVYPGEDIYACTEINFITLI 155

RESULT 6

US-08-705-660-34
; Sequence 4, Application US/08/705-660

; Patent No. 5858683

; GENERAL INFORMATION

; APPLICANT: KEESEE, SUSAN

; APPLICANT: OBAR, ROBERT

; APPLICANT: WU, YING-JYE

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE

; DETECTION OF CERVICAL CANCER

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Testa, Hurwitz & Thibault

; STREET: 125 High St.

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110

; COMPUTER AVAILABLE FROM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/705,660

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: GREENHALGH, DUNCAN A

; REGISTRATION NUMBER: 38,678

; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 248-7000

; TELEFAX: (617) 248-7100

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3224 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-705-660-34

Query Match 3.8%, Score 108.5, DB 2, Length 3224;

Best Local Similarity 22.1%, Pred. No. 0.7,
Matches 63, Conservative 17, Mismatches 86, Indels 94, Gaps 15;

QY 267 VNVASCL-----SFHKEDKTFHFVDEPKT-----KKEKSTFYFTALVL 305

```

Db      576 VHWAECLQKQSGLSNLSFYDQF-----YIGRSVHYWKKVLP LLLKLIKKNKSIPEPIDPLPK 631
Qy      306 YHH-----INAYEDGHVFDIVAYRDNLSYD-----MFYLLK 338
Db      642 HFHSVDIQASEIVEYBEDAHIITAILDAVNGNIEDAVTAFESIKSVVSYWNLALIFHRKA 691
Qy      339 LDKDFEVNKKLT--SIPTCKR-----FVPLQYUKDAEVSGLNVLKLP TSAIAVKKK 489
Db      692 --EDIE--NDALSPEEQEETKNYLKRTDYLIKIIDSDSNL--SVVKKLPVPLESVKEMLN 747
Qy      390 STYCOPEILCEG-----LPLPR-----VNYDYN--CKKYKYVYVATEVOWSPVP 430
Db      748 SVMGELEYEGGPLYKNGSLRNADSEIKRSTPSPTYSLSKSKYKSPKTPPRWAEDQ 807
Qy      431 TKIAKLVQKVEVL-----HWGEDHCWPSF---PIFVP 460
Db      808 NSLLKMKICQVFAIKKEMQELKLNSSNSASPHRWPTENYCPDSVP 852

```

RESULT 7

```

US-08-989-045-34
; Sequence 34, Application US/08989045
; Patent No. 6027905
; GENERAL INFORMATION:
; APPLICANT: KEESEE, SUSAN
; APPLICANT: OHAR, ROBERT
; APPLICANT: WU, YING-YE
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL CANCER
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,045
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GREENHALGH, JENNAN A
; REGISTRATION NUMBER: 38,678
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3224 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-989-045-34

```

```

Query Match      3.8%; Score 108.5; Db 3; Length 3224;
Best Local Similarity 22.1%; Fied. No. 0.7,
Matches 64; Conservative 37; Mismatches 86; Indels 99; Gaps 15;

Qy      267 VNNASCL-----SFHKEDKTFHFVDKRT-----KKEVSTKFTYDALVL 305
Db      576 VHWAECLQKQSGLSNLSFYDQF-----YIGRSVHYWKKVLP LLLKLIKKNKSIPEPIDPLPK 631
Qy      306 YHH-----INAYEDGHVFDIVAYRDNLSYD-----MFYLLK 338

```

```

Db      632 HFHSVDIQASEIVEYBEDAHIITAILDAVNGNIEDAVTAFESIKSVVSYWNLALIFHRKA 691
Qy      339 LDKDFEVNKKLT--SIPTCKR-----FVPLQYUKDAEVSGLNVLKLP TSAIAVKKK 489
Db      692 --EDIE--NDALSPEEQEETKNYLKRTDYLIKIIDSDSNL--SVVKKLPVPLESVKEMLN 747
Qy      390 STYCOPEILCEG-----LPLPR-----VNYDYN--CKKYKYVYVATEVOWSPVP 430
Db      748 SVMGELEYEGGPLYKNGSLRNADSEIKRSTPSPTYSLSKSKYKSPKTPPRWAEDQ 807
Qy      431 TKIAKLVQKVEVL-----HWGEDHCWPSF---PIFVP 460
Db      808 NSLLKMKICQVFAIKKEMQELKLNSSNSASPHRWPTENYCPDSVP 852

```

RESULT 8

```

US-09-134-078-61
; Sequence 61, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 08010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-134-078-61

```

```

Query Match      3.8%; Score 106; Db 4; Length 663;
Best Local Similarity 21.4%; Fied. No. 0.09;
Matches 133; Conservative 65; Mismatches 210; Indels 214; Gaps 40;

Qy      30 GVLLKNGFGMHITIGTKY-----NHWFD-----GLALLHSFTFKNGEYVYRSKYL 74
Db      23 GKFAI NCKEFPFPGSNYYMHYKSNCGMIDSVIFSAQDMGKVIPIWGFIDGFSYCRGK 80
Qy      75 PSTDYCNIPFANDIVVSFECTMAYPDPYKNIPAKAFSYLSHTI---PEPTUNCLINIMKT 141
Db      81 -----NTYMHPEPGVEGVEGISNA--QSGFERKIDYTVAKAKELGKIVIVVNN 128
Qy      132 GND-----YYATSFTNFIP--KIDPDQITFTIDKVDYSKYVA--VNIAISHPH 174

```


APPLICATION NUMBER: conversion to Provisional application)

```

; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: 115/09/458.791

```

APPLICATION NUMBER: US/09/458,791


```

; FILING DATE: 10-Dec-1999
; CLASSIFICATION : Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/958,598
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2631
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)470-4189
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1569 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-458-791-2

```

```

Query Match      3.5%; Score 100; DB 4; Length 1568,
Best Local Similarity 19.6%; Pred. No. 1.5;
Matches 71; Conservative 58; Mismatches 137; Indels 96; Gaps 15;

QY 122 ENCLINIMKT-----GDYYATSETNFIRKIDFQILETELEKVDYSKYVAVNLATS 171
DB 632 EQCPVAVEKISGGGPKPKENKGNKINQALQVFIKSIIEPKVSIIGK-----SNVIVT 683

QY 172 HPHYDSAGNI--LNGTISIVDKG--PTKYVL-----FKIPSSVPEKEKKSCFKHLEV 221
DB 684 GANFTRASNITMLKGTSTCDKDIQVSHVLNTHMKFSLPSS--RKEMKDVCIOFDGNG 741

QY 222 CS-----IPSRSLQFS-----YYHSGITERYIVFIEQFFKLDIVKLA 260
DB 742 CSSVGSLSYIALPHCSLIIPATTWISGGQNTMMGRNFDVIDNLIIISHELKGNINSEYC 801

QY 261 TAYIRGVNWSCLSFHKEDKTWTFHVDKTKKEVSTKFTYDALVLYHHIINAYEEDGHVVF 320
DB 802 VA-----TYCGFLAPSLK---SSKVRTNVTVKLRVQDTYLDGC--- 836

QY 321 DIVAYRDNLSLYDMFLKLDKDFEVNKKLTISPTCKRFVVPLOYDKDAEVSGLVKLPST 380
DB 837 -TLQYFEDPRTGYPVFS-----EVDTELE-----VKIKEND---NFNISKKDIE 878

QY 381 ATAVKEKGSIVCOPEILCEGIELPRVNYDNGKKYKYVATEVQNSPVPTKIKNVQT 440
DB 879 IILPHGNGQINQSFENITPPNQDTITTICTKIKGIKTASTIANSSE--KVEVKLGNELEY 936

QY 441 KE 442
DB 937 EQ 938

```

```

RESULT 13
US-09-459-066-2
; Sequence 2, Application US/09459066
; Patent No. 6187909
; GENERAL INFORMATION:
; APPLICANT: Spriggs, Melanie
; TITLE OF INVENTION: VIRAL-ENCODED SHMADPHOKIN PROTEIN
; TITLE OF INVENTION: RECEPTOR DNA AND POLYPEPTIDES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janis C. Henry
; STREET: 51 University St
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: MS-DOS/Windows 95
; SOFTWARE: WinF for Windows 95, 7.03
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09459,066
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/958,598
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2631
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)470-4189
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1568 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-459-066-2

```

```

Query Match      3.5%; Score 100; DB 4; Length 1568;
Best Local Similarity 19.6%; Pred. No. 1.5;
Matches 71; Conservative 58; Mismatches 137; Indels 96; Gaps 15;

QY 122 ENCLINIMKT-----GDYYATSETNFIRKIDFQILETELEKVDYSKYVAVNLATS 171
DB 632 EQCPVAVEKISGGGPKPKENKGNKINQALQVFIKSIIEPKVSIIGK-----SNVIVT 683

QY 172 HPHYDSAGNI--LNGTISIVDKG--PTKYVL-----FKIPSSVPEKEKKSCFKHLEV 221
DB 684 GANFTRASNITMLKGTSTCDKDIQVSHVLNTHMKFSLPSS--RKEMKDVCIOFDGNG 741

QY 222 CS-----IPSRSLQFS-----YYHSGITERYIVFIEQFFKLDIVKLA 260
DB 742 CSSVGSLSYIALPHCSLIIPATTWISGGQNTMMGRNFDVIDNLIIISHELKGNINSEYC 801

QY 261 TAYIRGVNWSCLSFHKEDKTWTFHVDKTKKEVSTKFTYDALVLYHHIINAYEEDGHVVF 320
DB 802 VA-----TYCGFLAPSLK---SSKVRTNVTVKLRVQDTYLDGC--- 836

QY 321 DIVAYRDNLSLYDMFLKLDKDFEVNKKLTISPTCKRFVVPLOYDKDAEVSGLVKLPST 380
DB 837 -TLQYFEDPRTGYPVFS-----EVDTELE-----VKIKEND---NFNISKKDIE 878

QY 381 ATAVKEKGSIVCOPEILCEGIELPRVNYDNGKKYKYVATEVQNSPVPTKIKNVQT 440
DB 879 IILPHGNGQINQSFENITPPNQDTITTICTKIKGIKTASTIANSSE--KVEVKLGNELEY 936

QY 441 KE 442
DB 937 EQ 938

```

```

RESULT 14
US-08-257-073-3
; Sequence 3, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA

```

```

1  ZIP: 15036
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: Patent in Release #1.0, Version #1 30
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/257,073
16
17 FILING DATE: 09-JUN-1994
18
19 CLASSIFICATION: 424
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: US 08/075,783
24
25 FILING DATE: 11-JUN-1993
26
27 PRIOR APPLICATION DATA:
28
29 APPLICATION NUMBER: US 07/852,305
30
31 FILING DATE: 18-MAR-1992
32
33 PRIOR APPLICATION DATA:
34
35 APPLICATION NUMBER: US 07/672,183
36
37 FILING DATE: 20-MAR-1991
38
39 ATTORNEY/AGENT INFORMATION:
40
41 NAME: Frommer, William S.
42
43 REGISTRATION NUMBER: 25,506
44
45 REFERENCE/DOCKET NUMBER: 454310-2570
46
47 TELECOMMUNICATION INFORMATION:
48
49 TELEPHONE: (212) 840-3333
50
51 TELEFAX: (212) 840-0712
52
53 TELEX: 425066 CURTMS
54
55 INFORMATION FOR SEQ ID NO: 3:
56
57 SEQUENCE CHARACTERISTICS:
58
59 LENGTH: 984 amino acids
60
61 TYPE: amino acid
62
63 STRANDEDNESS: Single
64
65 TOPOLOGY: linear
66
67 MOLECULE TYPE: peptide
68
69 FRAGMENT TYPE: internal
70
71 US-08-257 073 3

```

```

Query Match      3.5%; Score 98; DB 2; Length 984;
Best Local Similarity 19.1%; Pred. No. 1.1;
Matches 67; Conservative 49; Mismatches 104; Indels 130; Gaps 16;

125 LINIMKTGDY----YATSETFIR-KIIPQILLELA--KVDYSKYVAVNLAISHHYDYS 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
236 LGNTPTGKNFFLVVYKENTILLKWKYVGEKZGILLNNKVEVEKYZ 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 AGNILLNMGTSIVKGPTKYVLEKIPSSVPEKFKKSPFKHLEVCVSPSSQLQPSVYHS 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
286 -----INKE-----TPFTSILLHAYKKB 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 FGI-----TENYIVFIEQPKELDIVKLATAYPG-VNNWASCI SFHFEYTFWFEVDKPTK 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
305 NGTNILTESKNYALGSDIPKQDIL-ASNLFLSGNFNIEKFKQCAL--LVKFNKN 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
293 EVSTKFTYUALV-LYHIINAY----EEDGHVVFDIVAYRONSLYDMF----- 334
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 DVCKYKLSEDIVSKFEIKAEETDEDDEDDYETAYKLTESIDNLVKMKFTNNNDKSELIK 416
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
335 -----YLUKLD----KDFEVANKLISIPTEKFFVVPVLQYUKIAEFGS 372
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 LEEVDSSKLLENNYCSLLKQVDVDTGTLIDNYGCMENMDIFNNLKRLDI---YISEENINT 474
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 NLVKLPISATAVREKDGSIYCPGLCEGIELPRNYD-----VNGK 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
474 LNKKEPNAACVLKNVDVDT-----VNGKGLVPELNVYDLEVENHLYNDK 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
US-08-184-004-120
: Sequence 120, Application US/08184009
: Patent No. 5843975
: GENERAL INFORMATION:
: APPLICANT: Paoletti, Enzo
: APPLICANT: Turtaglia, James


```

QY 181 ILNMGTSLIVKGPRTKYVI FKIPSSVPEKPKKSCPKHI EVVCSIPSPSLLQPSYHSPGI 240
DB 181 ILNMGTSLIVKGRTKYVLFKIPSSVPEKPKKSCPKHLEVCSIPSPSLLQPSYHSPGI 240
QY 241 TENYIVFIEQPFKIDIVKLATAYIRGVNWSCLSPHKEDKTFHFVDRKTKKEVSTKPYT 300
DB 241 TENYIVFIEQPFKIDIVKLATAYIRGVNWSCLSPHKEDKTFHFVDRKTKKEVSTKPYT 300
QY 301 DALVLYHHINAYEDGHVDFDI VAYRNSLYDMFYLLKKLDKDFEVNKKLTSTPTCKRPFV 360
DB 301 DALVLYHHINAYEDGHVDFDI VAYRNSLYDMFYLLKKLDKDFEVNKKLTSTPTCKRPFV 360
QY 361 PLOYKDAEVSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKYYK 420
DB 361 PLOYKDAEVSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKYYK 420
QY 421 ATEVOMSPVPTKIAKLVOTKEVLHMGEDHCWSEPIFVPSDAREDEGVVLTGVVSE 480
DB 421 ATEVOMSPVPTKIAKLVOTKEVLHMGEDHCWSEPIFVPSDAREDEGVVLTGVVSE 480
QY 481 PNKAPFLLILDAKTFKELGRATVNVEMHLDLHGMFIPONDGAETE 526
DB 481 PNKAPFLLILDAKTFKELGRATVNVEMHLDLHGMFIPONDGAETE 526

```

RESULT 2

```

US-10-053-192-4
: Sequence 4, Application US/10053192
: Publication No. US20030087336A1
: GENERAL INFORMATION:
: APPLICANT: BACHMANN, Heinrich
: APPLICANT: BRUGGER, Roland
: APPLICANT: FRIEDLEIN, Arno M
: APPLICANT: WIRTZ, Gabriele M
: APPLICANT: WOGGON, Wolf-Dietrich
: APPLICANT: WYSS, Adrian
: APPLICANT: WYSS, Markus
: TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
: FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...
: CURRENT APPLICATION NUMBER: US/10/053,192
: PRIOR FILING DATE: 2002-01-15
: PRIOR APPLICATION NUMBER: 103382.0
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 4
: LENGTH: 506
: TYPE: PRT
: ORGANISM: CHICKEN
US-10-053-192-4

```

```

Query Match
Best Local Similarity 96.3%; Score 2721; DB 5; Length 506,
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 EEHPEPIKAEVQGLPTWLOGVLLRNGPGMHTIGDTKYNHWFPGALHSHFTFKNGEVY 69
DB 1 EEHPEPIKAEVQGLPTWLOGVLLRNGPGMHTIGDTKYNHWFPGALHSHFTFKNGEVY 60
QY 70 RSKYLRSDTYNKNIEANPIVVSFPG--TMAYDPCKNIFAKAFYSLSHTIPFTDNCILINIM 129
DB 61 RSKYLRSDTYNKNIFANPIVVSFPGTMAYDPCKNIFAKAFYSLSHTIPFTDNCILINIM 120
QY 140 KTGDYATSETNFIKIDPQTLTLDDKVDYSKYVAVNLATSPHYDSAGNINMGTSIV 189
DB 121 KTGDYATSETNFIKIDPQTLTLDDKVDYSKYVAVNLATSPHYDSAGNINMGTSIV 180
QY 190 UKGRTKYVLFKIPSSVPEKPKKSCPKHLEVCSIPSPSLLQPSYHSPGITENYIVFIE 249
DB 181 UKGRTKYVLFKIPSSVPEKPKKSCPKHLEVCSIPSPSLLQPSYHSPGITENYIVFIE 240

```

```

QY 250 QPPKLPDIVKLATAYIRGVNWSCLSPHKEDKTFHFVDRKTKKEVSTKPYTDAI VIYHHI 409
DB 241 QPPKLPDIVKLATAYIRGVNWSCLSPHKEDKTFHFVDRKTKKEVSTKPYTDAI VIYHHI 400
QY 310 NAYEEDGHVDFDI VAYRNSLYDMFYLLKKLDKDFEVNKKLTSTPTCKRFPVPIQYIKDAE 469
DB 301 NAYEEDGHVDFDI VAYRNSLYDMFYLLKKLDKDFEVNKKLTSTPTCKRFPVPIQYIKDAE 460
QY 370 VGSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKYYKYYVATEVOMSPV 429
DB 361 VGSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKYYKYYVATEVOMSPV 420
QY 430 PTKIAKLVOTKEVLHMGEDHCWSEPIFVPSDAREDEGVVLTGVVSEFNKAIFILL 489
DB 421 PTKIAKLVOTKEVLHMGEDHCWSEPIFVPSDAREDEGVVLTGVVSEFNKAIFILL 480
QY 490 LDKTFKELGRATVNVEMHLDLHGMF 515
DB 481 LDKTFKELGRATVNVEMHLDLHGMF 506

```

RESULT 3

```

US-10-053-192-5
: Sequence 5, Application US/10053192
: Publication No. US20030087336A1
: GENERAL INFORMATION:
: APPLICANT: BACHMANN, Heinrich
: APPLICANT: BRUGGER, Roland
: APPLICANT: FRIEDLEIN, Arno M
: APPLICANT: WIRTZ, Gabriele M
: APPLICANT: WOGGON, Wolf-Dietrich
: APPLICANT: WYSS, Adrian
: APPLICANT: WYSS, Markus
: TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
: FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...
: CURRENT APPLICATION NUMBER: US/10/053,192
: PRIOR FILING DATE: 2002-01-15
: PRIOR APPLICATION NUMBER: 103382.0
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 5
: LENGTH: 529
: TYPE: PRT
: ORGANISM: BOVINE
US-10-053-192-5

```

```

Query Match
Best Local Similarity 35.9%; Score 1014.5; DB 9; Length 529;
Matches 226; Conservative 9; Mismatches 184; Indels 44; Gaps 20;
QY 10 EEHPEPIKAEVQGLPTWLOGVLLR--NGPGMHTIGDTKYNHWFPGALHSHFTFKNGEV 67
DB 1 EELSSPLTAHVGTGRIPLWLTGSLRKTGTFGLFEVSGSEPPYHLFDGALHSHFTFKNGEV 60
QY 68 YRSKYLRSDTYNKNIEANPIVVSFPG--TMAYDPCKNIFAKAFYSLSHTIPFTDNCIL 125
DB 61 TYHPRFRTDQYVPMTEKPIVITREKGTTCAPPDCKNIFSPFFSYFPGV--EVDNAL 118
QY 126 INIMKTDDYATSETNFIKIDPQTLTL--UKVDYSKYVAVNLATSPHYDSAGNIN 183
DB 119 VNYVPGEDYATSETNFIKIDPQTLTL--UKVDYSKYVAVNLATSPHYDSAGNIN 178
QY 184 MGTSTVUKGRTKYVLFKIPSSVPEKPKKSCPKHLEVCSIPSPSLLQPSYHSPGITEN 243
DB 179 IGCFCGKNFSIANIVKIPLOADKELPSKETS--EIVVQFPCSRPFKYSYHSPGITEN 247
QY 244 YIVFIEGPFKIDIVKLATAY--HGVNWSCLSPHKEDKTFHFVDRKTKKEVSTKPYT 400
DB 238 YIVFIEGPFKIDIVKLATAY--HGVNWSCLSPHKEDKTFHFVDRKTKKEVSTKPYT 297
QY 301 DALVLYHHINAYFETSHVVFPIVAYFET--VLYKKIITKPF--VNNKLTSTPT 454

```

Db 298 SPFNLFPHINTYEDNGFLVDCCWKGFEVYNYFTLYCANRENEEYKKNARKAPOE 357
QY 355 CKRVPVLOYDKDAEVSNLVKLP-TSATAV--KEKDGSIYCOPEILCEG---IFLPRV 407
Db 358 VRYVPLNIDK-ADTKNLVTLPTNTATALLCSDEFTTWLEPFVLFSSPQAEFFQI 416
QY 408 NYD-YNGKKYKYVATEVQWSPVTKIAKLNVOTKEVLH--WGEDHCWPSSEDFIVPSDA 464
Db 417 NYQYCKPYTYAYGLGNHF-VPRICKLNVRKTKFTWFTVQWPPDSYSPRPIYSHDA 475
QY 465 REEDGGVLTCTGVVSEPNKAF-ELLILDAKTFKELGRA--TVNVEHMLDGHMF 515
Db 476 LEEDGGVLSVSVSPGAGKPAYLLILNAKOLSEVARAEFTVEINIPVTFHGLF 529

RESULT 4

US-09-758-269-18

; Sequence 18, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: UCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 18

; LENGTH: 538

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-18

Query Match 11.08; Score 311; DB 10; Length 538;
Best Local Similarity 23.88; Pred. No. 1.3e-16;
Matches 135; Conservative 78; Mismatches 213; Indels 144; Caps 25;

QY 9 KEEHPEIK-AEVGGQLPTWGLVLLRNGSGMHTIGDTKYN-----HWF-DGLALLHSPTF 62
Db 56 RDETP-PVKDLPVHGFLPECLNGEFVRVGP-----NPKFDVAVGHWFDGDMHGVRI 108
QY 63 KNGVYYSKYLSRSTYCNIEANPIVVSFFGTMAYPDCKNIFAKAFSYLSHTIPEFTD 122
Db 109 KDGKATVVSRY-----VKTSLKQEEF-----FGAAKFMKIGDLKGFFG 147
QY 123 NCLINI-----MKTGDYYATSETN-----FIKIDPQTLTILD 156
Db 148 LLWNVVLLKTLKLLNIVGNGJANIALVYHGGKLLALQEAQKIVYFVIFSDQLIG 207
QY 157 KVDYSKYVAVNLATSHPHYDS-AGNILNMGTSLVKGPKIKYVLKIPSSVPEKEKKSCF 215
Db 208 IIDYDKRLTHSF-TAHPKVPDVTGEMTFGYS-----HTPPYLTFRVIS-----KUGI 254
QY 216 KHLWCVCSIPSRLQSYYSHSFGITENYIVTFEQFKL---DIVKLATAYIKGVNWSG 272
Db 255 MHDVPVITIS-----EPIMHDFAITETVAIFMDLPMHFRPKEMVK-----EKKMI 300
QY 273 LSEHKEDKTFHFVDRKTKKEVSTK-FYTDALVLYHHINAYEEDGHVVFIVAYRDSNLY 331
Db 301 YSPDPTKARFGLPRYAKDELIRWFLPNCFLPHNANAWEEDEVL-ITCRLENPDL 359
QY 332 DMFYLLKLDKDFEVNNKLSIPTCKRFVVPVLDKDAEVSGLVKLPTSATAVKKQDSI 391
Db 360 DMVSGKVEKLENFGNELIYEM---RF-----NKKTQSA 389

QY 352 YLDFEILCEGLFELPFWNYFNKSKYKVVYATAFVWSEVPKIFAKINVOI----- 440
Db 390 -SOKKLSASAVDPRIECYTGKKQYVYGTILDSIAKVTGIIKFDLHAEAFIGKRMLEK 448
QY 441 ---FEVLHWFETHTWPSFFIVFVPSFARFEFEVVI-TVVVFEFNAFELIITAFIFE 496
Db 449 GGNTEKGIYDGGSPR-YGSFAIYVPPPTA-FEDDGYLIFPVHNFNTGKS-CVIVTIATKMS 505
QY 497 ELGRATVNV--EMHLDLHGMFIFQNDLGAET 525
Db 506 AEPVAVVELPHRVPGFHALFVTEBOLEOT 536

RESULT 5

US-09-758-269-8

; Sequence 8, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: UCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; CURRENT FILING DATE: 2001-01-12

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 8

; LENGTH: 538

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-8

Query Match 11.08; Score 310; DB 10; Length 538;
Best Local Similarity 23.68; Pred. No. 1.3e-16;
Matches 135; Conservative 79; Mismatches 213; Indels 144; Caps 25;

QY 9 KEEHPEIK-AEVGGQLPTWGLVLLRNGSGMHTIGDTKYN-----HWF-DGLALLHSPTF 62
Db 56 RDETP-PVKDLPVHGFLPECLNGEFVRVGP-----NPKFDVAVGHWFDGDMHGVRI 108
QY 63 KNGVYYSKYLSRSTYCNIEANPIVVSFFGTMAYPDCKNIFAKAFSYLSHTIPEFTD 122
Db 109 KDGKATVVSRY-----VKTSLKQEEF-----FGAAKFMKIGDLKGFFG 147
QY 123 NCLINI-----MKTGDYYATSETN-----FIKIDPQTLTILD 156
Db 148 LLWNVVLLKTLKLLNIVGNGJANIALVYHGGKLLALQEAQKIVYFVIFSDQLIG 207
QY 157 KVDYSKYVAVNLATSHPHYDS-AGNILNMGTSLVKGPKIKYVLKIPSSVPEKEKKSCF 215
Db 208 IIDYDKRLTHSF-TAHPKVPDVTGEMTFGYS-----HTPPYLTFRVIS-----KUGI 254
QY 216 KHLWCVCSIPSRLQSYYSHSFGITENYIVTFEQFKL---DIVKLATAYIKGVNWSG 272
Db 255 MHDVPVITIS-----EPIMHDFAITETVAIFMDLPMHFRPKEMVK-----EKKMI 300
QY 273 LSEHKEDKTFHFVDRKTKKEVSTK-FYTDALVLYHHINAYEEDGHVVFIVAYRDSNLY 331
Db 301 YSPDPTKARFGLPRYAKDELIRWFLPNCFLPHNANAWEEDEVL-ITCRLENPDL 359
QY 332 DMFYLLKLDKDFEVNNKLSIPTCKRFVVPVLDKDAEVSGLVKLPTSATAVKKQDSI 391
Db 360 DMVSGKVEKLENFGNELIYEM---RF-----NKKTQSA 389
QY 392 YCQPEILCEGLFELPFWNYFNKSKYKVVYATAFVWSEVPKIFAKINVOI----- 440
Db 440 -SOKKLSASAVDPRIECYTGKKQYVYGTILDSIAKVTGIIKFDLHAEAFIGKRMLEK 448

Db 575 LKLL--EATVKLPSRVPGYGFHGTFFINANDL 601

RESULT 8

US-09-758-269-10

; Sequence 10, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09758.269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 10

; LENGTH: 577

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-10

Query Match 9.3%; Score 263; DB 10; Length 577;

Best Local Similarity 23.8%; Pred. No. 8.2e-13;

Matches 129; Conservative 97; Mismatches 189; Indels 128; Gaps 32;

QY 19 EVQGLPTWGLVLLRNG--PGMHIGDTKYNHWFGLALLH--SETFKNGEYVYRSKYL 74

Db 118 EVVGIPSCLGKGVIRNCANMPFPPLAG--HHLFPGDGMTHAVSIGFDN-QVYSYSCPT 173

QY 75 RSDTYNCNLEARNIV-VSEFTMAYPDCKNI-----FAKAFSYLSHTIPEFTDNCI-IN 127

Db 174 RT-----NRIVQETALGRSVEPKPIGELHGHSLARLALFTAPAGIGLVGDRGMG 224

QY 128 IMKTGDDYV-----ATSEIN--FIRKIDPO--TLETL-----DKVDYSKYVAVNATSHP 173

Db 225 VANAGVVFENGKLLAMSEDDLPYOVKIDQCDDLETIGFF:FDQI:ISS-----VIAHP 277

QY 174 HYD-SAGNII NMGTSTIVUKGRTKYVLEKIPSSVPEKEKKSCFKHL-EVWCSTPSLSLQ 231

Db 278 KVDATTGDLHTLSYNVLKPKHLRYLKF-----NTCGKTRDVEITLP-----E 320

QY 232 PSYHSFGITENYIVFIHQF--FKLDIVKLATAYIKGVNWSCLSHHKEUKTWFRFVDKK 289

Db 321 PTMHDFAITENFVVPQVQVVKL-----SEMIRG---GSPVIVYVKKKMAFGVLSKQ 371

QY 290 --TKKEVSTKFTYDALVLYHHINAYE---EDGHVVFIDIVAYRONSLYDMFYLLKLDKDFE 344

Db 372 DLTGSDINWVDPICF-CFHLNNAWEETEGDPVIVIG-----SCMSPPDTIFS 421

QY 345 VNNKLTSLTECFKFFVVPQVQVVKL-----SEMIRG---GSPVIVYVKKKMAFGVLSKQ 404

Db 422 ESGRTFREL -----SEIRLNM-----PTKESN-----PKVIVIGVNL 454

QY 405 P--RVNDYNGKKKYVYATEVQVSPVTKIAKLNVTKEVLHWGDEHCWPS-----EPIF 458

Db 455 EACHINRSYVRGKSGQVVIATADPWPKSGIAKVDIQNGIV---SEFNYSRFGEGPCF 511

QY 459 VVSPDAREDEGVLTGVVSEPNKAPFLILDAKTEKELGRATVNVEMHLDLHGMFIPO 518

Db 512 VPEGES-FEDGCVMGFVHUEKESEF-VVVLATENKQVAVALPEKVPYGFHGTFFVSE 569

QY 519 NDL 521

Db 570 NDL 572

RESULT 9

US-09-758-269-6

; Sequence 6, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09758.269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 599

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-6

Query Match 9.2%; Score 260; DB 10; Length 599;

Best Local Similarity 24.4%; Pred. No. 1.5e-12;

Matches 133; Conservative 85; Mismatches 217; Indels 110; Gaps 29;

QY 11 EHPEVKAEEVQULPILWGLVLLRNGPK-MHTIGDTKYNHWFGLALLHSEFTKNGEYVY 69

Db 132 EGVKPNIPVVKIPKIPSTKGVYVFNGANPLHE--PVISHHEFFITGCMVHAVKRFHRSASY 140

QY 70 RSKYLPSTDYNCNTEANRIVVS-EFGTMAYPDCKNI-----FAKAFSYLSHTIPEFTDNCI 123

Db 130 ACFF-----TQINKEVQEKUGKVFVKAKAGELHGHGTGAPLMLFYAFAAGIVDP 240

QY 124 CL-INIMTIGENY-----ATSEIN--FIRKIDPO--TLETLKVVYVSKYVAVNATSHPH 174

Db 241 ABETGVANAGLVYFNKFLAMSEDDLPYQVQVVKL-----SEMIRG---GSPVIVYVKKKMAFGVLSKQ 299

QY 175 YD-SAGNII NMGTSTIVUKGRTKYVLEKIPSSVPEKPKKSCFKHI-FVWCSTPSLSLQPS 243

Db 300 VVPS--GELFALSYVWVSKPVLYKYPFFS-----PSTKSPD-----VFI-----QIDQPT 344

QY 234 YVHSFGITENYIVFIHQF--FKLDIVKLATAYIKGVNWSCLSHHKEUKTWFRFVDPKTK 291

Db 344 MMHDFAITENFVVPQVQVVKL-----SEMIRG---GSPVIVYVKKKMAFGVLSKQ 394

QY 292 KEVSTFTYDA IVLVHHINAYE PGHVVEIVAYVSENSLYEMFYLLKLDKDFEVN 346

Db 395 DSSNIK-WIDAPDCFCFHLNNAWEETEGDPVIVIG-----SCMSPPDTIFS 442

QY 347 NKLTSLTECFKFFVVPQVQVVKL-----SEMIRG---GSPVIVYVKKKMAFGVLSKQ 402

Db 443 ENKLSVL-----SEIRLNL-----KTGSTERPILSNDQVNL 476

QY 403 ELPRVNDYNGKKKYVYATEVQVSPVTKIAKLNVTKEVLHWGDEHCWPSSEPIFVPS 461

Db 477 EAGVNVNANLGGKTKFKAYLALAEADPWKVSCHAKVLLJIGFVKKHIYGINPYGGEPLFPG 536

QY 452 PIAPKFFSGVLT--VVVSEPNKAPFLILDAKTEKELGRATVNVEMHLDLHGMFIPO 519

Db 537 -EGGDEEGYIL-CFVIDEKTWKSELOIVNVSLEV--EATVKLPSRVPGYGFHGTFFIGAD 592

QY 520 PLGAE 524

Db 593 DLAKO 597

RESULT 10

US-08-976-063C-22

Db 477 EVGMVNRLLGSHRYAYLAVAPWPKESGFARKEIDSTGELTKFYCGCH-FCGRCPCFVP 535
QY 461 SPDA-----REDEGVVLTVCVVVSEPNKAPFLILLDAKTFKELGRATVNVEMHLLDLMGIT 516
Db 536 MDPAARHPRGDDGVLT-FVHDERAGTSELLVYNAADIRLEATVGLPSRVPFGHGTFI 594
QY 517 PONDIGAE 524
Db 595 FQLEFAQ 602

RESULT 12

US-09-758-269-4

; Sequence 4, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2000-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 4

; LENGTH: 595

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-4

Query Match

Best Local Similarity 8.3%; Score 233.5; DB 10; Length 595;

Matches 119; Conservative 86; Mismatches 210; Indels 135; Gaps 24;

QY 20 VQGLPTWLOGVLLRNGPGMHTIGDTKYNHWFQGLALLHSTFKNGEVIYRSYLRSDIY 79
Db 130 IHTLPLSLNGAYTRNGPNQFPLRGPY-HLFDGDMHLAIKTHNGKATLCRSYVK--TY 186
QY 80 NCNIE-----ANRWSEFGTMAYPDPCKNIFAKATS 111
Db 187 KYNVEKOTGAPVMPNVESFGNGVTASVARGALTAARVLTQY-----NPVNGI----- 234
QY 112 YLSHTIPEFTDNCINIMKTGGDYATSENF---IPKIDPQITFLDKVDYKYVAVNL 168
CL 235 GLANTSHAFFSNRL-----FALGESDLFYAVELTESGTEIGRYDEGLKLSM 284
QY 169 ATSHPHYDS-AGNILNMOTSIVDKGRTKYVLFKIPSSVPEKEKKKSCFKHLEVVCISPSR 227
Db 285 -TAHPTKTDPIGETFAFRYGVPPFLT-YFEP---DSAGKQR-----DVPIF 327
QY 228 SLQPSYVHSFGITENYIVF--IEQPFKLDIVKLATAYIRGVNWSCLSHFKDKTWPHF 285
Db 328 SMTSPSELDRAITKRIHAIFAELIQLGRMNMILDIV-----LEGSPVGTQDKCTPLGV 381
QY 286 VDRKTKKEVSTK-FYTDALVYHHINAY-PEDGHVVDIVAYRDNLSYDMFYLLKLDKDF 343
Db 382 IPKYAGDESEMKWFEPVGFNIHAINAWDEDDGN---SVVLIAPNIMSTIEHTLERMD--- 435
QY 344 EVNNKLTISPTCKFVVPLOQYDKDAEVSNIKLPTSATAVEKDGSIYQCEILCEGIE 403
Db 436 -----LVHALKEVKVLDIVTGIVRRHP-ISARNLD 464
QY 404 LPRVNYNGKKYKYVYATEVQWSPVPTKIATLNVQ-----TKVLIHMGDHCWPFSEP 456
Db 465 FAVINPAFLGRCSYVYAAIGDPMPKISGVVVKLVNSKGRDNDCTVAPPMYGSQ-QYCGRP 523

QY 457 IFV---PSPDAREDEGVVLTVCVVVSEPNKAPFLILLDAKTFKELGRATVNVEMHLL-IDL 511
Db 534 FFVARDFGNFAEEDDGVVTVYHDEVTVGESKP-LVWAKSKSELEFIVAVALPFRVYGF 582
QY 512 HGMETPQNDL 521
Db 583 HGLFVKESDI 592

RESULT 13

US-09-758-269-12

; Sequence 12, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 12

; LENGTH: 612

; TYPE: PRT

; ORGANISM: Vigna unguiculata

US-09-758-269-12

Query Match

Best Local Similarity 7.8%; Score 220; DB 10; Length 612;

Matches 126; Conservative 98; Mismatches 215; Indels 136; Gaps 30;

QY 1 METINRNKKEHPPIKAE-----VQGLPTWLOGVLLKNG-- 36
Db 116 VETALVSHPEKHPILPKTADPPVQIASNFAVPVPEHAAQGLPVVVKIPKCIDGVVVRNGAN 175
QY 37 PSMHTIGDTKYNHWFQGLALLHSTFKNGEVIYRSYLRSDIYFANPIVVSHPGTM 96
Db 176 PLYEPVAG---HFEEDGVMHVAVKFTNGAASACRF--TFTQPI SQF-----KSLGRP 224
QY 97 AYDPCKNI---FAKAFSVLSHTIPEFTDNCIL-INIMKTGGDY-----ATSENF-- 143
Db 225 VPKAIGELHGHSGIAPILLPYACIPLCLVWGSGGCMVANAGLVYFNNHLLAMSEDLUPY 284
QY 144 -IPKIDP-TLETLKVIYSKYVAVNLAISHPHYISA-INTJNM-TSIVIKSFKYVLFKI 201
Db 285 HVPITPNQDITVGPYDFNFGINSTM-IAHPRILPVPVPLDHAISYNVILKPYLYKFRFS- 342
QY 202 FSSV PEKKEKKSCFKHLEVVCISIPSPSLQTSYVHSFGITENYIVFIEQF--EELIV 257
Db 343 PDGVKSPDVE-----IP---LKEPTMHDFAITENFVVVPPQVQVFKL--- 382
QY 258 KIATAYTEGVNWSCLSHFKELKTWFHFLVKFKTKVESTREYIDA--LVLYHHINAYEE- 314
Db 383 ---TEMTIG---GSPVVVDKNTSPFGIL-HKNAKDANAMRWIDAPICPCFHLNNAWEP 435
QY 315 --DGHVVDIVAYRDNLSYDMFYLLKLDKDFEVENKLTISPTKPFVPLQYVKKDAEVS 372
Db 436 ETEEVVIGSGMTPADSIFN-----EGEESLSKVL-----SEIRL 470
QY 373 NLVKLPTSATAVEKDGSIYQCEILCF---CIELPRVNYNGKKYKYVYATEVQWSP 428
Db 471 NL-----PTKSTSTPPP-IISDAEQVNIAGMNPKNLGRTOGFAYLALAEPP 517
QY 429 VPTKIATLNVQKEV--LHWTFHCFWPSFTFVESPIAPPEFEGSVLITVWVSEPNFAP 486
Db 518 KVSQFAKVDLLSGHVKKMYGCEKA-FGSHPLI--PWGJGJEDDYSILAFVHHEKWKSE- 573

QY 487 LLLLDKATFKELGKATVNVEMHLDLHGMFIPQNDL 521
 DB 574 LOIVNAONKLEASIKLPSRPVPGFHGTF:HSKDL 608

RESULT 14
 US-09-758-269-2
 : Sequence 2, Application US/09758269
 : Patent No. US20020104120A1
 : GENERAL INFORMATION:
 : APPLICANT: TUCHI, SATOSHI
 : APPLICANT: KOBAYASHI, MASATOMI
 : APPLICANT: SHINOZAKI, KAZUO
 : TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
 : FILE REFERENCE: 3914-3
 : CURRENT APPLICATION NUMBER: US/09758269
 : PRIOR FILING DATE: 2001-01-12
 : PRIOR FILING DATE: 2001-01-11
 : PRIOR APPLICATION NUMBER: JP 2000-010056
 : PRIOR FILING DATE: 2000-01-13
 : SOFTWARE: Patent In Ver. 2.1
 : SEQ ID NO 2
 : LENGTH: 583
 : TYPE: PRT
 : ORGANISM: Arabidopsis thaliana
 : OTHER INFORMATION: Unknown Organism
 : US-09-758-269-2

Query Match 7.3%, Score 207, DB 10, Length 583,
 Best Local Similarity 20.2%, Pred. No. 2.6e-08;
 Matches 111; Conservative 51; Mismatches 206; Indels 142; Gaps 26;

QY 20 VQOGLPTWIQVLLPNSPGMHITGDTKYNNWFDGLALLHSFTFKNGEVYRKYLRSDTY 79
 DB 124 VEGTIPGIDGVYLRNGAN-PMFEPTAGHHLPDGDGMVHAKITNGSASYACRFKTE-- 180
 QY 80 NCNTEANKIVVSE-FGTMYAPDCKNI-----FAKAFS----- 111
 DB 181 -----RLVQEKPLGPPVPKPAIGFTHGHSGLIAPLMI FYAPRLGGL INNQGVGVANAG 233
 QY 112 --YLSHTIPETUNCL---INIMKTGMDYATSEINFIRKIDQILLETLDKVDYKYVAV 166
 DB 234 LVYNNKLLAMSEDDLPYQLKITQTD-----LOTVGRYDFDQGLK- 274
 QY 167 NLATSHPHYDSAGNIN-MGJISIVGKGFKYVLFKIPSSWPKEKKKSCFKHLEVCSIP 225
 DB 275 SAMLAHPKLDIVTKELHAIISYDVVKKPKLYKFFES-PDGKSPS-----LEI----- 320
 QY 226 SPSTIQPSYHSEGTITNYIVFIQPP--FKI-DIVKLATAYIPGVANWASCLSFHKEDKTW 282
 DB 321 --PLETPTMIDHUALIENFVV1PDQOVVFKLGEMISGKSPVVFDEGKVSRLGIMPKDAT- 377
 QY 283 PBFVDRTKKEVSTKFTYDA--LVLYHHINAYEEDGHVVDIVAYRDNLSYDMFYLLKLD 340
 DB 378 -----BASQIIWNSPETCFCHLNNAWES----- 401
 QY 341 KDFEVNKKITSTPTCKRFPVPLGVGDKIAEVSGLNVLKPLTSATAVKEDGSLVYQUPILCE 400
 DB 402 ---PETEEIVVIGSCMSPAUSIFNERDESLKSVLSIRINLRTKTRTSLLINFDV--- 455
 QY 401 GIELPRVNDYNGKKYKYV-ATEVQMSVPVTKIAKLNQV---KEVLHMGEDHWPSP 456
 DB 456 NLEIGMVRNRKLGPKTFPAFLAIPW-PKVSFGAKVDLCTGHMKKYIYGGEK--YGGFP 512
 QY 457 IFVPSDA---REDEGVLITCVVSPHPKAPPELLILDAKTEKELCBATVNVEMHL--DL 511
 DB 511 FFLPNSNGNEEDGGYTF-CHVDEETKSELQIINAVNLK--EATIKLPSRPVPGF 569
 QY 512 HCMFIPQNDL 521

DB 570 HGTFVDSNEL 579

RESULT 15

US-09-884-465A-338
 : Sequence 338, Application US/09884465A
 : Publication No. US20030077293A1
 : GENERAL INFORMATION:
 : APPLICANT: Shire Biochem, Inc.
 : APPLICANT: Hamel, Jusee
 : APPLICANT: Brodeur, Bernard
 : APPLICANT: Martin, Denis
 : APPLICANT: Charland, Nathalie
 : APPLICANT: Ouellet, Catherine
 : TITLE OF INVENTION: Streptococcus Antigens
 : FILE REFERENCE: 055190-0044
 : CURRENT APPLICATION NUMBER: US/09884465A
 : PRIOR FILING DATE: 2001-06-20
 : PRIOR APPLICATION NUMBER: 60,212,683
 : PRIOR FILING DATE: 2000-06-20
 : NUMBER OF SEQ ID NOS: 384
 : SOFTWARE: Patent In version 3.1
 : SEQ ID NO 338
 : LENGTH: 888
 : TYPE: PRT
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: Unknown Organism
 : US-09-884-465A-338

Query Match 3.8%, Score 108; DB 9; Length 888;

Best Local Similarity 21.4%, Pred. No. 3.8;

Matches 135; Conservative 76; Mismatches 245; Indels 174; Gaps 32;

QY 9 KEEHPEPIKAEVQGLPTWLGQVLLNPSGPMH-TIGDTKYNNWFTGSLAL-IHSFTFKNGE 66
 DB 258 KFSAPATGCPQIGQ-PILFNNSLATPSLPIPGTSHKKEGXYGPDANKILAEKKS 416
 QY 67 VYSPSKYLPSTNYNFIANFIVVSPST-----MAYPDP-KNIFAKAFASYLS 114
 DB 317 GFVNSYEFKKLLIEQLKAAKHLLEVKTSNGLDLSLSSHEQVYFQ-N--AKEMKOLD 471
 QY 115 HTIPEFTDNCINIMKTGDDYYATSEINFIRKIDPOTLETLDKVDYKYVAVNLSHHP 174
 DB 372 KKIEE---KIAGIMK---QYGVKRESIVVNKKNALTYPADPIDEHK-PWGIGHSHSN 422
 QY 175 YF-----SAGNTIINMG---TSIVD--KQPI-----KVLEKLESSVPE 207
 DB 423 YELEKPEGVAKKGGNKVYTGHELTNVVNLKKNSTNNONFTLANGKKKVSFSP-PE 474
 QY 208 KKKKSKFKHLEVCSIPSSILQPSYHSEGTITNYIVFIE-----QPERLDIVKL 259
 DB 480 LEKKLGI--NMLVFLITPI--GFVLEKVS--KVFGEV--NIANFELI--PVL--P--FYVLIASK 547
 QY 260 ---ATAVIPGVNWSCLSFHKEDKTWPH-FVDPKTKKEVSTKPY-----TDALVLY--HH 408
 DB 538 DYPEVSYDGTFTVPTSLAYKMASOTIYPFPHAGDTYLRVNPQFAPVAPKGTDALVRVDFPH 597
 QY 309 INAYEEDGH-----VVEDIVAYRDN----- 428
 DB 598 GNAYLENNYKVGELKLPKLNQGTTRTAGNKTPVTFMANAYLDNOSTYIVAVPILLENEN 657
 QY 329 -----SLYDMFYKLIKUNDFEVVANKITSTPTKFPFVVPVLCYMDKAPVQ--SNVVKLP 478
 DB 658 QTKPSTILQPKFKPKAGKSKLDEKVEEPTSEK-----VEKEKLSGTGNSNSTLEKVP 714
 QY 379 TSATAVKEDGST---YCOPEILCEGIELPNNVNIYN-KKYVYVATFVQWSIVPTKIAK 435
 DB 714 T-VDPQVEKVAKFAESY-----GKLENVLFNMGITFIYLSGQV--IFKNMAD 760
 QY 436 LNVQTEKVLHMGEDHWPSPSEPIFVPSPDAREDEGVLLTCVVVSEFNK-APFLLILDART 494
 DB 761 FTGEAPQ---GNGENKFS-----NGKVSIGTIVENQPIENKFIALSLPFAFN 803



GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on July 15, 2003, 09:36:25, Search time: 04 seconds
(without alignments)
1290 247 Million cell updates/sec

Title: US-10-053-192-1

Perfect score: 526

Sequence: 1 METIFNRKNKEHPPIKAEV... MHLIQLHMFIFLNIHAEIE 526

Scoring table: Q1150
Gapop 60 0, Gapext 60.0

Searched: 671580 seqs, 205047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_prote.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	526	100.0	526	13	Q91994
2	25	4.8	547	4	Q9HAY6
3	25	4.8	547	4	Q9NVH5
4	25	4.8	566	11	Q9LNS6
5	25	4.8	566	11	Q9ERN9
6	25	4.8	566	11	Q9LX15
7	13	2.5	516	13	Q90WH4
8	11	2.1	420	5	Q9KKW9
9	11	2.1	620	5	Q9VFS2
10	10	1.9	533	13	Q9VFX2
11	10	1.9	549	13	Q90WH3
12	8	1.5	164	2	Q9AEG4
13	8	1.5	287	14	Q9PFA2
14	8	1.5	293	11	Q9CTP3
15	8	1.5	333	5	Q9CFP5
16	8	1.5	483	11	Q9VHP2

17	8	1.5	532	11	Q9QNF1
18	8	1.5	533	4	Q16518
19	8	1.5	533	5	Q28175
20	8	1.5	533	6	Q9TVB8
21	8	1.5	533	6	Q97623
22	8	1.5	533	6	Q9XT71
23	8	1.5	533	6	Q05661
24	8	1.5	533	11	Q9VX76
25	8	1.5	533	11	Q91Z05
26	8	1.5	533	13	Q9VY12
27	8	1.5	539	4	Q96JY5
28	8	1.5	545	4	Q96J08
29	8	1.5	550	4	Q9BYV7
30	8	1.5	554	14	Q9KUS4
31	8	1.5	580	2	Q9JPJ5
32	8	1.5	580	16	Q93Y22
33	8	1.5	580	15	Q93Q01
34	8	1.5	724	10	Q9LKT8
35	8	1.5	856	11	Q62121
36	8	1.5	876	11	Q60752
37	8	1.5	1011	5	Q9Y148
38	8	1.5	1059	5	Q9VIX6
39	7	1.3	53	3	Q9TGT9
40	7	1.3	78	16	Q92JB4
41	7	1.3	92	16	Q8R9N7
42	7	1.3	105	2	Q54742
43	7	1.3	111	17	Q9Y8Y6
44	7	1.3	124	2	Q05206
45	7	1.3	124	10	Q49708

ALIGNMENTS

RESULT 1

Q91993	PRELIMINARY:	PRT:	526 AA.
AC	Q91993		
DI	01-OCT-2000 (IREMBLrel. 15, Created)		
LI	01-OCT-2000 (IREMBLrel. 15, Last sequence update)		
PT	01-DEC-2001 (IREMBLrel. 19, Last annotation update)		
DE	Beta carotene 15,15'-dioxygenase (EC 1.13.11.21).		
GN	BCDO.		
OS	Gallus gallus (Chicken).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
KC	STRAIN=LSL LOHMANN; TISSUE=DUODENUM;		
PX	MEDLINE=20261261; PubMed=14799297;		
RA	Wyss A., Wirtz G.M., Woggon W.D., Bruggen R., Wyss M., Friedlein A.,		
RA	Bachmann H., Hunziker W.:		
FT	"Cloning and expression of beta,beta carotene-15,15'-dioxygenase";		
FL	Biochem Biophys Res Commun 271:344-346(2000).		
RN	[2]		
PP	SEQUENCE FROM N.A.		
KC	STRAIN=LSL LOHMANN; TISSUE=DUODENUM;		
PX	MEDLINE=21134366; PubMed=11237856;		
PA	Wyss A., Wirtz G.M., Woggon W.D., Bruggen P., Wyss M., Friedlein A.,		
RA	Kiss G., Bachmann H., Hunziker W.:		
FT	"Expression pattern and localization of beta,beta-carotene 15,15'-		
PT	dioxygenase in different tissues";		
FL	Biochem J 354:531-539(2001).		
DR	EMBL AJ271386, CAB92827.1, ..		
DR	InterPro: IPR004294; PPF65.		
FW	Prim: PF04055; PPF65; 1.		
FW	Dioxygenase, Cofactor:beta.		
SQ	SEQUENCE 526 AA: 67484 MW: 83074AAAP75FE3458 CRC64;		

Query Match: 100.0%, Score: 526, LS: 13, Length: 526;
Best Local Similarity: 100.0%, Pred No: 0;

```

Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METFNKNKEHPPEIKAEVQGLPTWLCQVILRNGPGMHTIGDTKYNHWFGLALLHSF 60
DB 1 METFNKNKEHPPEIKAEVQGLPTWLCQVILRNGPGMHTIGDTKYNHWFGLALLHSF 60

QY 61 TFKNCEVYRSKYLRSOTYNCNIFANKIVVSEFGIMAYDPCKNIFAKAFSYLSHTIPEF 120
DB 61 TFKNCEVYRSKYLRSOTYNCNIFANKIVVSEFGIMAYDPCKNIFAKAFSYLSHTIPEF 120

QY 121 TDNCLINIMKTGDYYATSETNFIRKIDPOTLETLDKVDYSKYVAVNLTATSHPHYDSAGN 180
DB 121 TDNCLINIMKTGDYYATSETNFIRKIDPOTLETLDKVDYSKYVAVNLTATSHPHYDSAGN 180

QY 181 ILNMGTSIVDKGRTKYVLFKIPSSVPEKPKKSCFKHLEVCVCSIPSRSLQPSYHSFGI 240
DB 181 ILNMGTSIVDKGRTKYVLFKIPSSVPEKPKKSCFKHLEVCVCSIPSRSLQPSYHSFGI 240

QY 241 TENYIVFIEQPFKLDIVKLATAYIRGVNMAASCLSPHKEDKTFHFVDRKTKKEVSTKEYT 300
DB 241 TENYIVFIEQPFKLDIVKLATAYIRGVNMAASCLSPHKEDKTFHFVDRKTKKEVSTKEYT 300

QY 301 DALVLYHHINAYEDGHVDFDIWAYRDNLSYDMFYLLKLDKDFEVNKNKLTSTPTCKPFV 360
DB 301 DALVLYHHINAYEDGHVDFDIWAYRDNLSYDMFYLLKLDKDFEVNKNKLTSTPTCKPFV 360

QY 461 PLOYDKDAEFGVSNIVKLTPTSATAVPEFDFSTYQGFILFEGFIPFPVNYDYNCKKYKYY 420
DB 461 PLOYDKDAEFGVSNIVKLTPTSATAVPEFDFSTYQGFILFEGFIPFPVNYDYNCKKYKYY 420

QY 421 ATEVQWSPVTKIAKLVQTKVELVHWGDEHCWSPSEIFVPSDPAREDFGVVLTCTVWVSE 480
DB 421 ATEVQWSPVTKIAKLVQTKVELVHWGDEHCWSPSEIFVPSDPAREDFGVVLTCTVWVSE 480

QY 481 PNKAPFILLIATKTEKELGPATVNVEMHLLGLHGMFIPQNDLGAETE 526
DB 481 PNKAPFILLIATKTEKELGPATVNVEMHLLGLHGMFIPQNDLGAETE 526

RESULT 2
Q9HAY6
ID Q9HAY6 PRELIMINARY: PRT; 547 AA.
AC Q9HAY6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).
GN BCD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan W., Jang G.-P., Haesleer F., Esumi N., Chang J., Kerrigan M.,
RA Campochiaro M., Campochiaro P., Palczewski K., Zack D.J.;
RT "Cloning and characterization of a human beta, beta-carotene 15,15'-
RT dioxygenase that is highly expressed in the retinal pigment
RT epithelium."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases
DR EMBL: AF294900; AAG15380.1; -
DR InterPro: IPR004294; RPE65.
DR Pfam: PF04055; RPE65; 1.
KW Dioxygenase; Oxidoreductase.
SQ SEQUENCE: 547 AA; 62637 MW; F94BC8B01056F9CB CRC64;

Query Match 4.8%; Score 25; DB 4; Length 547;
Best Local Similarity 100.0%; Pred. No. 20-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 NIEANRIVVSEFGTMAYPDPCKNIF 106
DB 82 NIEANRIVVSEFGTMAYPDPCKNIF 106

RESULT 3
Q9NVH5
ID Q9NVH5 PRELIMINARY: PRT; 547 AA.
AC Q9NVH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ10730 f1s, clone NTZHP3001307.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wadatsuma M., Hosogiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project."
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001592; BAA91776.1; -
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.
SQ SPQENFE 547 AA; 62574 MW; 494A1400105606FF CRC64;

Query Match 4.8%; Score 25; DB 4; Length 547;
Best Local Similarity 100.0%; Pred. No. 20-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 NIEANRIVVSEFGTMAYPDPCKNIF 106
DB 82 NIEANRIVVSEFGTMAYPDPCKNIF 106

RESULT 4
Q9JJS6
ID Q9JJS6 PRELIMINARY: PRT; 566 AA.
AC Q9JJS6;
DT 01-MAR-2001 (TrEMBLrel. 15, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).
GN RCD01 OR RCD0 OR RCD0.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA SPAIN-C57BL; TISSUE=KIDNEY;
RA Wyss A., Witz G.M., Woggon W.D., Brugger K., Wyss M., Friedlein A.,
RA Bachmann H., Hunziker W.;
RT "Expression pattern and localization of beta, beta-carotene 15,15'-
RT dioxygenase in different tissues."
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA R-Jmd1 T.M., Grotjeman S., Duncan T., Yu S., Wiggert K., Gantt E.,
RA Cunningham F.X.;
RT "Identification, expression and substrate specificity of a mammalian
RT beta-carotene 15,15'-dioxygenase."
RL J. Biol. Chem. 0:0-0(2001);
DR EMBL: AJ278564; CAB92521.2; -
DR EMBL: AF271298; AAG3982.1; -
DR MGI:1926923; Bcd01.
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.

```

KW Dioxxygenase; Oxidoreductase.
 SQ SEQUENCE 566 AA; 63864 MW; 1B4367815247A9C2 37674.
 Query Match 4.8%, Score 25, DB 11, Length 566,
 Best Local Similarity 100.0%, Pred. No. 2.1e-18;
 Matches 25, Conservative 0, Mismatches 0, Indels 0, Gaps 0;
 QY 82 NIEANRIVVSEFGTMAYDPCKNIF 106
 Db 82 NIEANRIVVSEFGTMAYDPCKNIF 106
 RESULT 5
 Q9ERN9
 ID Q9ERN9 PRELIMINARY: PRT: 566 AA
 AC Q9ERN9
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE beta, beta-carotene 15,15'-dioxxygenase (EC 1.13.11.21).
 GN BCD01.0P RCPD.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL; TISSUE=KIDNEY;
 RA Yan W., Jiang G.-F., Hasegawa F., Esumi N., Chang J., Korrigan M.,
 RA Campochiaro M., Campochiaro P., Palczewski K., Zack D.J.;
 FT "Cloning and characterization of a human beta, beta carotene 15,15'-
 RT dioxxygenase that is highly expressed in the retinal pigment
 RI epithelium."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases
 DR EMBL: AF294899; AAC15381.1;
 DR MGD: MGI:1926923; Bcd01.
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65, 1.
 KW Dioxxygenase; Oxidoreductase.
 SQ SEQUENCE 566 AA; 63852 MW; 2C4D7719C456790 78044.

Query Match 4.8%, Score 25, DB 11, Length 566,
 Best Local Similarity 100.0%, Pred. No. 2.1e-18;
 Matches 25, Conservative 0, Mismatches 0, Indels 0, Gaps 0;
 QY 82 NIEANRIVVSEFGTMAYDPCKNIF 106
 Db 82 NIEANRIVVSEFGTMAYDPCKNIF 106
 RESULT 6
 Q91XT5
 ID Q91XT5 PRELIMINARY: PRT: 566 AA.
 AC Q91XT5
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE beta-carotene 15,15'-dioxxygenase.
 OS Pampus norvegicus (Pai)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTES;
 RA Takitani K., Ban R., Tamai H.;
 FT "Regulation of beta-carotene 15,15'-dioxxygenase in oxidative stress".
 PL Submitted (JUN 2001) to the EMBL/GenBank/DBJ databases
 DR EMBL: AB062912; BAB60807.1;
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65, 1.
 KW Dioxxygenase.
 SQ SEQUENCE 566 AA; 63637 MW; A1FFB47BA6CE6E5 0P064;

Query Match 4.8%, Score 25, DB 11, Length 566,
 Best Local Similarity 100.0%, Pred. No. 2.1e-18;
 Matches 25, Conservative 0, Mismatches 0, Indels 0, Gaps 0;
 QY 82 NIEANRIVVSEFGTMAYDPCKNIF 106
 Db 82 NIEANRIVVSEFGTMAYDPCKNIF 106
 RESULT 7
 Q90WH4
 ID Q90WH4 PRELIMINARY: PRT: 516 AA.
 AC Q90WH4
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Putative b,b-carotene-15,15'-dioxxygenase.
 GN B-DIOX.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=21216714; PubMed=11278918;
 RA Kiefer C., Hessel S., Lampert M., Vogt K., Lederer M.O.,
 RA Breithaupt D.B., von Lintig J.;
 FT "Identification and characterization of a mammalian enzyme catalyzing
 RT the asymmetric oxidative cleavage of provitamin A";
 RL J. Biol. Chem. 276:14110-14116(2001).
 RP EMBL: AJ290590; CA247566.1;
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65, 1.
 KW Dioxxygenase.
 SQ SEQUENCE 516 AA; 58584 MW; DEUA47707E55C93F C8C64;

Query Match 4.5%, Score 13, DB 13, Length 516;
 Matches 13, Conservative 0, Mismatches 0, Indels 0, Gaps 0;
 QY 94 GTMAYDPCKNIF 106
 Db 94 GTMAYDPCKNIF 106
 RESULT 8
 Q9NKG9
 ID Q9NKG9 PRELIMINARY: PRT: 620 AA.
 AC Q9NKG9
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE DRPE65.
 GN RPE65 OR RPE65 OR Q39347.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Aphidoidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-P;
 RA Sahara H., Suzuki E.;
 FT "Drosophila cDNA similar to RPE65".
 PL Submitted (Apr 2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB041507; BAA94508.1;
 DR FlyBase: FB0038171; RPE65.
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65, 1.
 KW Dioxxygenase.
 SQ SEQUENCE 620 AA; 66047 MW; 4F4C245C19CB3424 C8C64;

Query Match 2.18; Score 11; DB 5; Length 620;
 Best Local Similarity 100.0%; Pred. No. 0.0083;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 453 PSEPIFVPSPD 463

DB 550 PSEPIFVPSPD 560

RESULT 9

Q9VFS2 PRELIMINARY; PRT; 620 AA.

AC Q9VFS2;

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DE CG9447 protein (15,15', beta carotene dioxigenase)

GN RPE65 OR BETA-DIOX OR CG9447

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota, Metazoa, Arthropoda, Insecta;

OC Elysiptoda, Neoptera, Eudopterygota, Lepidea, Brachycera, Muscomorpha;

OC Ephydroidea, Drosophilidae; Drosophila

OX NCBI_TaxID=7227;

PP SEQUENCE FROM N.A.

RA STRAIN=HEKLEY;

RX MEDLINE=20196006; PubMed=10731132.

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P., Hoskins P.A., Galie P.F.,

RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin E.,

RA Ballou K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.T., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Hurlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra T.,

RA Cherry J.M., Cawley S., Dahlve C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dudan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glisok A., Gong F., Grifelli J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jabali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodita C.D., Kraft C., Kravitz S., Kulp P., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy R., Murphy L., Muzey D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,

RA Palazzolo M., Piltman G.S., Fan S., Pollard T., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders P.P.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos L., Simpson M., Skupski M.P., Smith T.,

RA Spier E.C., Spradling A., Stappleton M., Sprong P., Sun F.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z. Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Winchester T., Wright K.C., Wu D., Yand S., Yan Q.A.,

RA Ye J., Yeh R.-F., Zaveri I.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zhang X.H., Zhang F.N., Zhang W., Zhao X., Zhu S., Zhu X., Smith H.C.,

RA Gibbs K.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000)

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=20224792; PubMed=10766819;

RA von Lintig J., Vogt K.;

RT "Filling the gap in vitamin A research."

RL J. Biol. Chem. 275:11915-11923(2000)

DR EMBL; AF003701; AAF54978.1; -

DR EMBL; AJ276682; CAB93141.1; -
 DR Flybase: FBgn0038171; RPE65.
 DP InterPro: IPR004294; RPE65.
 DP Pfam: PF03055; RPE65.1.
 KW Dioxigenase.
 SQ SEQUENCE 520 AA; 69341 MW; 91E722E8162C31C3 CRC64;

Query Match 2.18; Score 11; DB 5; Length 620;
 Best Local Similarity 100.0%; Pred. No. 0.0083;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 453 PSEPIFVPSPD 463

DB 550 PSEPIFVPSPD 560

RESULT 10

Q9YGX2 PRELIMINARY; PRT; 533 AA.

AC Q9YGX2;

DT 01-MAY-1999 (TREMblrel. 10, Created)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DE RPE65 protein.

DE RPE65 protein.

GN RPE65.

OS Gallus gallus (Chicken).

OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;

OC Archosauria, Aves, Neognathae, Galliformes; Phasianidae;

OC Gallus.

OX NCBI_TaxID=9031;

PP SEQUENCE FROM N.A.

RA Hirosewa K., Sagar H.;

RA "Investigation of an endoplasmic reticulum related protein in the

RA vertebrate retinal pigment epithelium."

RA Thesis (1998), Institute of Medical Science, The University of Tokyo.

RA EMBL; AB017594; BAA75667.1; -

DR InterPro: IPR004294; RPE65.

DP Pfam: PF03055; RPE65.1.

DR SEQUENCE 533 AA; 60930 MW; 67E6A5B5993084A0 CRC64;

Query Match 1.98; Score 10; DB 13; Length 533;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 97 AYDPCKNIF 106

DB 107 AYDPCKNIF 116

RESULT 11

Q90WH3 PRELIMINARY; PRT; 549 AA.

AC Q90WH3;

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DE Putative b,b'-carotene-9',10'-dioxigenase.

DE Putative b,b'-carotene-9',10'-dioxigenase.

GN B-DIOX-II.

OS Eukaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;

OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

PX MEDLINE=2121714; PubMed=11778918;

RA Kiefer C., Hessel S., Lampert J.M., Vogt K., Lederer M.O.,

RA Breithaupt D.E., von Lintig J.;

RT "Identification and characterization of a mammalian enzyme catalyzing

RT the asymmetric oxidative cleavage of provitamin A."

DR EMBL; AF003701; AAF54978.1; -


```

R1 J Biol. Chem. 275:14110-14116(2001).
DR EMBL: AJ290391; CAC37567.1; -.
DR InterPro: IPR004294; RPF65.
DR Pfam: PF03055; RPF65.1.
KW Dioxigenase.
SQ SEQUENCE 549 AA; 62426 MW; 6B4FCAD1BF87DBR7 CRC64;

Query Match 1.9%; Score 10; DB 13; Length 549;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NRIVVSFRPT 95
DB 111 NRIVVSFRGT 120

RESULT 12
Q9AEG4 PRELIMINARY; PRT; 169 AA.
AC Q9AEG4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Putative AcpA protein (Fragment).
GN AcpA.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=548;
RN [1] SEQUENCE FROM N.A.
RP STRAIN-BW16627;
RC "The AcrA/AcrB/ToIC efflux pump participates in multidrug resistance
RT in Enterobacter aerogenes."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DP EMBL: AJ306389; CAC35722.1; -.
FT NON_TER 169
SQ SEQUENCE 169 AA; 19014 MW; 33EBC25007JFD0E5 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 OTLETLDK 157
DB 67 OTLETLDK 74

RESULT 13
Q9RJA2 PRELIMINARY; PRT; 287 AA.
AC Q9RJA2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative binding protein dependent transport protein.
GN Scd552 or ScF91.22.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1] SEQUENCE FROM N.A.
RP STRAIN-A3(2);
RC Oliver K., Harris D.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases
RN [2] SEQUENCE FROM N.A.
RP STRAIN-A3(2);
RC Thomson N.R., Parkhill J., Barrrell J.G., Rajandream M.A.;
PA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE-97000351; PubMed=8841436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RN Mol. Microbiol. 21:77-96(1996).
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Clater K.F., Cardeno-Liarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrrell J.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RN Nature 417:141-147(2002).
DR EMBL: AL132973; CAB61179.1; -.
DR InterPro: IPR000515; BPD.transp.
DR Pfam: PF00528; BPD.transp.1
SQ SEQUENCE 287 AA; 31533 MW; 77AB4EDEC0FFA6CF CRC64;

Query Match 1.5%; Score 8; DB 16; Length 287;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 LGRATVNV 505
DB 14 LGRATVNV 21

RESULT 14
Q9CTP3 PRELIMINARY; PRT; 293 AA.
AC Q9CTP3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE A930029L06Rik protein (Fragment).
GN A930029L06Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1] SEQUENCE FROM N.A.
RP STRAIN-C57BL/6J; TISSUE=RETINA;
RX MEDLINE-zl085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gajjala I., Bodo H., Kasukawa I., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gasterland F., Gissi C., King K., Kochiwa H.,
RA Kuchi P., Lewis S., Matsuo Y., Nakado I., Resole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gastonich S., Hill D., Hofmann M., Hume E.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., McBeris P.,
RA Nordone P., Ping R., Ringwald M., Rodriguez J., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
PA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
PA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."

```

RL Nature 409:685-690(2001).
 DR EMBL: AK020906; BAB32248.1; -.
 DR MCD: MGI:192506R; A90029106Rik.
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 FT NON_TER 293 293
 SQ SEQUENCE 293 AA; 33281 MW; F4B867CDEF4A3E7D CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DR 11; Length 293;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PDPCKNIF 106
 |||||
 Db 109 PDPCKNIF 116

RESULT 15

Q95RP9 PRELIMINARY; PRT; 333 AA.
 AC G95RP9;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE LD16758p.
 GN GGT OR PCTNA-GH04245 OR C610392.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Aqbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George P.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Gelniker S.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY061225; AA128773.1; -.
 DR Flybase: FBgn0040295; ogt.
 SQ SEQUENCE 333 AA; 37253 MW; 9c75223E506D9BF9 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DR 5; Length 333;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 KIDPQTL 153
 |||||
 Db 146 KIDPQTL 153

Search completed: July 15, 2003, 09:42:43
 Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 21:28:47, Search time: 787 seconds
(without alignments)
11502.837 Million coll updates/sec
Title: US-10-053-192-2
Perfect score: 3111
Sequence: 1 cggatccacagtaacggcc aaaaaaaaaaaaaaaaaaaaaa 3111

Scoring table: OLIGO-NUC

Gapop 60.0, Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4109,780

Minimum DB seq length: 0

Maximum DB seq length: 200,000,000

Post-processing: Listing first 45 summaries

Database: GenEmbi:

1: gb_fa.*
2: gb_hg.*
3: gb_in.*
4: gb_lm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_b4.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_mv.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_dli.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_ror.*
36: em_htg_wam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_higo_hum.*
40: em_higo_mus.*
41: em_htgo_other.*

Pred No is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3111	100.0	3111	5	GA271386
2	3111	100.0	3111	5	AX0271386
3	3111	100.0	3111	5	AX0271386
4	36	1.2	219	6	AX112509
5	36	1.2	219	6	AX112509
6	36	1.2	219	6	AX112509
7	36	1.2	219	6	AX112509
8	36	1.2	219	6	AX112509
9	36	1.2	219	6	AX112509
10	36	1.2	219	6	AX112509
11	36	1.2	219	6	AX112509
12	36	1.2	219	6	AX112509
13	36	1.2	219	6	AX112509
14	36	1.2	219	6	AX112509
15	36	1.2	219	6	AX112509
16	36	1.2	219	6	AX112509
17	36	1.2	219	6	AX112509
18	36	1.2	219	6	AX112509
19	36	1.2	219	6	AX112509
20	36	1.2	219	6	AX112509
21	36	1.2	219	6	AX112509
22	36	1.2	219	6	AX112509
23	36	1.2	219	6	AX112509
24	36	1.2	219	6	AX112509
25	36	1.2	219	6	AX112509
26	36	1.2	219	6	AX112509
27	36	1.2	219	6	AX112509
28	36	1.2	219	6	AX112509
29	36	1.2	219	6	AX112509
30	36	1.2	219	6	AX112509
31	36	1.2	219	6	AX112509
32	36	1.2	219	6	AX112509
33	36	1.2	219	6	AX112509
34	36	1.2	219	6	AX112509
35	36	1.2	219	6	AX112509
36	36	1.2	219	6	AX112509
37	36	1.2	219	6	AX112509
38	36	1.2	219	6	AX112509
39	36	1.2	219	6	AX112509
40	36	1.2	219	6	AX112509
41	36	1.2	219	6	AX112509
42	36	1.2	219	6	AX112509
43	36	1.2	219	6	AX112509
44	36	1.2	219	6	AX112509
45	36	1.2	219	6	AX112509

ALIGNMENTS

RESULT 1
LOCUS GA271386
DEFINITION Gallus gallus mRNA for beta-carotene 15,15'-dioxygenase (BCDO gene).
ACCESSION AJ271386
VERSION AJ271386.1
KEYWORDS BCDO gene; beta-carotene 15,15'-dioxygenase.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3111)
AUTHORS Wyss, A., Wirtz, G., Wengert, W., Progg, P., Wyss, M., Friedlein, A.,


```

/codon_line="regressor"
/tissue_type="colon carcinoma"
/codon_line="oligo-dt-primed cDNA in pcDNA3"
123..1070
/codon_start=1
/product="EGP-314 protein homologue"
/protein_id="CAA04496.1"
/db_xref="GI:2764595"
/db_xref="SPIRHH0:055159"
/translation="MAPPKALAFGIIIAVVTATIAAGKQVQNNYKILTSYFNFNQ
ECQCTSYQNTVIGSKLASKGLVMKAEIMSKSGRMKPEGAIONDGLYDPECDEQ
GLFKACQCATCMTAGVPTDKDEITCSEPTVYIIIEKHKAQAPYNF
SLTALQDTFASRYMLNPKFIKIMYENNVITIDLMONSOKTODDDVADVAYFEK
DVKGSLEFSSKSMOLRVANGELLDIDPGGILLYYDEKAPESMGLTAGIIVVV
VLAVIAGTVLVITIKRSKYEKAFIKEMGEIHKELNA"
BASE COUNT 417 a 339 c 411 g 357 t
ORIGIN
Query Match 1.2% Score 36. DB 10. Length 1524;
Best Local Similarity 100.0%; Pred. No. 7.3e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGCGCGCAGTGGTGGTGAAT 36
|||||
DB 18 CGGATCCACTAGTAAAGGCGCGCAGTGGTGGTGAAT 53
|||||

RESULT 8
AF073473
LOCUS Solanum tuberosum 1730 bp mRNA linear PLN 18-JUL-1998
DEFINITION Solanum tuberosum phospholipase C kinase precursor, mRNA, nucleotide
VERSION AF073473
KEYWORDS Solanum tuberosum; Solanaceae; Solanum
SOURCE Thorbjornsen, T.
ORGANISM Solanum tuberosum
REFERENCE 1 (bases 1 to 1730)
AUTHORS Thorbjornsen, T.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1998) Department of Plant Biology, Royal
Veterinary and Agricultural University, Thorvaldsensvej 40,
Frederiksberg C 1871, Denmark
FEATURES
source 1..1730
/organism="Solanum tuberosum"
/db_xref="taxon:4113"
/tissue_type="leaf"
209..1657
/codon_start=1
/product="phospholipase C kinase precursor"
/db_xref="GI:3328122"
/translation="MASATASHITFCGIPITSSSTTNSTTHSSARFLKTPRLRQF
AGAVAPLEINVAIKRLSKASSNPVAVSMARKSVGLSSDLKGGKVFVRADLN
VPLDQAQNTDITRIIPAAIPTIKHILANGAKVILSSHLGRKGVTPKYSIAPLVPLS
ELLGIQVADKICIGPEVKLVDSLEGGVLLLENVREYKEEKNEPEAKKLASLAD
LFVNDATGTAHAAHATGCVTKFKPSVAGFLKQKELDYLVAVSTPKPPAAIVGVS
KVSKSIGVTESLKEDKIDILLGGGMIFTFYKAGLGVSGSLVEEDKFLATSLLEKAK
AKGVSLIPSDVVIAUKPAPDANSKIPVASAIPDQWMDIDIGPDSIKTFNDALDTKT
VINQPMGVFEKATGTEATKRLADLSKGVITIIIGGDSVAAEKVGVSVMSSH
ISTGACSLLEKGVKLPGVIALDADAPAV"
BASE COUNT 449 a 400 c 406 g 475 t
ORIGIN
Query Match 1.2% Score 36; DB 8; Length 1730;
Best Local Similarity 100.0%; Pred. No. 7.2e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGCGCGCAGTGGTGGTGAAT 36
|||||

```

```

|||||
29 CGGATCCACTAGTAAAGGCGCGCAGTGGTGGTGAAT 64
|||||

RESULT 9
AF079853
LOCUS Arabidopsis thaliana 1982 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5965792.
VERSION AF079853
KEYWORDS Arabidopsis thaliana; Arabidopsis; Arabidopsis
SOURCE Unknown.
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 1982)
AUTHORS Schroeder, J.I., Antosiewicz, D.M., Schachtman, D.P. and Clemens, S.
TITLE Nucleic acids encoding metal uptake transporters and their uses
JOURNAL Patent: US 5965792-A 1 12-OCT-1999;
FEATURES
source 1..1982
/organism="unknown"
BASE COUNT 354 a 611 c 569 g 448 t
ORIGIN
Query Match 1.2% Score 36; DB 6; Length 1982;
Best Local Similarity 100.0%; Pred. No. 7.2e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGCGCGCAGTGGTGGTGAAT 36
|||||
DB 18 CGGATCCACTAGTAAAGGCGCGCAGTGGTGGTGAAT 53
|||||

RESULT 10
AF015523
LOCUS Arabidopsis thaliana 1984 bp mRNA linear PLN 29-SEP-1999
DEFINITION Arabidopsis thaliana low-affinity cation transporter (LCT1) mRNA,
complete cds.
VERSION AF015523
KEYWORDS Arabidopsis thaliana; Arabidopsis; Arabidopsis
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 1984)
AUTHORS Schachtman, D.P., Kumar, R., Schroeder, J.I. and Marsh, E.L.
TITLE Molecular and functional characterization of a novel low affinity
cation transporter (LCT1) in higher plants
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (20), 11079-11084 (1997)
MEDLINE 98021496
PUBMED 9380762
REFERENCE 2 (bases 1 to 1984)
AUTHORS Schachtman, D.P., Kumar, R., Schroeder, J.I. and Marsh, E.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1997) Botany, University of Adelaide, Adelaide,
SA 5005, Australia
FEATURES
source 1..1984
/organism="Arabidopsis thaliana"
/cultivar="Atlas 66"
/db_xref="taxon:4545"
/tissue_type="root"
/notes="Cloned from cDNA library by yeast complementation"
1..1984
/gene="LCT1"
137..1861
/gene="LCT1"
/codon_start=1
/product="low-affinity cation transporter"
/protein_id="AAR71851.1"
/db_xref="tr:2460044"

```



```

RESULT 14
ARI57785      ARI57785      870 bp  DNA      linear      PA: 17-OCI-2001
DEFINITION   Sequence 1 from patent US 6245898.
ACCESSION    ARI57785
VERSION      ARI57785.1 GI:16218795
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 870)
AUTHORS     Testa, J.K., Quigley, J.P. and Seandel, M.
TITLE       Monoclonal antibodies that recognize antigens associated with tumor
            metastasis
JOURNAL      Patent: US 6245898-A 1 12-JUN 2001;
FEATURES     location/qualifiers
            source
BASE COUNT   168 a 253 c 262 g 187 t
ORIGIN

Query Match      1.1%; Score 33; DB 6; Length 870;
Best Local Similarity 100.0%; Pred. No. 3.5e-06;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 ATCCACTAGTAAAGCGCGGCAGTGTGGTGGAAAT 36
        |||||||
Db      1 ATCCACTAGTAAAGCGCGGCAGTGTGGTGGAAAT 33

RESULT 15
AXI75601/C
Locus        AXI75601      275 bp  DNA      linear      PAT 03-JUL-2001
DEFINITION   Sequence 57 from Patent WO0144512.
ACCESSION    AXI75601
VERSION      AXI75601.1 GI:14598921
KEYWORDS     .
SOURCE       Norway Rat.
ORGANISM     Rattus norvegicus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE    1 (bases 1 to 275)
AUTHORS     Gould-Rothberg, B.E.
TITLE       Method of identifying ligands for the peroxisome proliferator
            activated receptor gamma using differential gene expression
JOURNAL      Patent: WO 0144512-A 57 21-JUN-2001;
            Curagen Corporation (US)
FEATURES     location/qualifiers
            source
BASE COUNT     62 a 74 c 79 g 57 t 3 others
ORIGIN

Query Match      1.0%; Score 30; DB 6; Length 275;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 30; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1 CGGATCCACTAGTAAAGCGCGGCAGTGTGG 30
        |||||||
Db      169 CGGATCCACTAGTAAAGCGCGGCAGTGTGG 140

```

Search completed: July 16, 2003, 02:26:10
 Job time : 7875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM nucleic - nucleic search, using sw model

Pub. on: July 15, 2003, 23:55:37 ; Search time 4167 seconds
(without alignments)
12091233 Milliseq call updates/sec

```

Title: US-10-053-192-2
Perfect score: 3111
Sequence: 1 cgaatccactactaagccc...

```

Scoring table: OLIG₀_Nuc

Searched. 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum	DR	seq	length
Maximum	pk	seq	length

post-processing: Listing first 45 summaries

```

Database .
FST.*
1: em_cstba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_ysst:*
18: em_ysst_hum,*
19: em_ysst_inv,*
20: em_ysst_fm,*
21: em_ysst_vrt,*
22: em_ysst_fun,*
23: em_ysst_man,*
24: em_ysst_mus,*
25: em_ysst_other,*
26: em_ysst_pro,*
27: em_ysst_ref,*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being plotted, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	586	18.8	661	14	HL89779 pgplc.pk0
2	35	1.2	73	9	AL449734 AL449734
3	36	1.2	197	9	AJ133823 AJ133823
4	36	1.2	259	14	BM87697 TM886.Hum
5	36	1.2	386	9	AA514191 HFLEST-74
C	36	1.2	418	14	W91597 MTA.G05.077

7	36	1,2	334	17	A2251210	mqM182P_R
8	36	1,2	451	9	A1449722	A1449722
9	36	1,2	452	10	BE414749	WML002_A0
10	36	1,2	458	13	B1293079	MC3MR_APP
11	36	1,2	496	17	A2251292	mqQ117P_R
12	36	1,2	503	9	AA551459	W514190_HFLEST-74
13	36	1,2	555	14	W91680	W91680_MTA_C02.079
14	36	1,2	612	14	RM889119	PM285459_TM077_Hum
15	36	1,2	633	14	BM887889	BM887889_TMT105_Hu
16	36	1,2	694	10	AW636749	AW636749_Gr-13_Glob
17	33	1,1	464	10	BE414750	WML002_A0
18	33	1,1	475	10	BE414889	WML001_G1
19	32	1,1	465	10	BE414893	WML001_H0
20	31	1,0	472	14	RM888080	TMT1504_Hu
21	31	1,0	471	10	BE414879	WML001_F0
22	31	1,0	467	10	BE414892	WML001_H0
23	30	1,0	285	14	W91767	W91767_MTA_B08_085
24	30	1,0	768	14	HQ797960	HQ797960_EST_5818
25	29	0,9	102	10	AW608948	PM3-PT004
26	29	0,9	103	12	RP927285	CM1_CT026
27	29	0,9	106	10	AW608921	PM3-PT004
28	29	0,9	108	10	AW608904	PM3-PT004
29	29	0,9	109	10	AW937014	PM3-DT004
30	29	0,9	110	10	AW608949	PM3-PT004
31	29	0,9	110	12	RP286599	RP286599_EST451290
32	29	0,9	111	10	AW608997	PM3-PT004
33	29	0,9	111	10	AW879277	PM3-CT001
34	29	0,9	111	12	RP286784	EST451375
35	29	0,9	113	10	AW608996	PM3-PT004
36	29	0,9	114	10	AW577820	CM0-CN000
37	29	0,9	115	10	AW647009	PM3-DT003
38	29	0,9	117	10	AW608943	PM3-PT004
39	29	0,9	119	10	AW608988	PM3-PT004
40	29	0,9	121	10	AW608918	PM3-PT004
41	29	0,9	123	10	AW608904	PM3-PT004
42	29	0,9	124	10	AW608903	PM3-PT004
43	29	0,9	124	12	RP927285	CM1_CT026
44	29	0,9	125	10	AW936837	PM3-DT002
45	29	0,9	126	10	AW947006	PM3-DT003

ALIGNMENTS

LOCUS R1389779 661 bp mRNA linear EST (n=Aug-2001)
DEFINITION popl.p002.al2 Primary Chicken Pituitary Hypothalamus/Pineal
Library Gallus gallus cDNA clone popl.p002.al2.5' similar to
cctb|CA00825.1 (A127186) beta-carotene 15,15'-dioxygenase (Gallus
gallus), mRNA sequence.
ACCESSION BI389779
VERSION R1389779.1 GI:15083061
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Cniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 661)
Porter, T. E. and Coqburn, I. A.
ESTs from Primary Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IFAFS Animal Genome Project
Unpublished (2001)
Contact: Larry A. Coqburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: coqburn@udel.edu, www.chickest.udel.edu.
Location/Qualifiers
1..661
/organism="Gallus gallus"

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 36; Conservative 0; Mismatches 0; Gaps 0;

QY 1 CGGATTCACCTAGTAAACGGCGGCGGAGTGTGTGTGGAAT 36
 |||
 DB 423 CGGATTCACCTAGTAAACGGCGGCGGAGTGTGTGTGGAAT 288

RESULT 7
 LOCUS A2254240 434 bp DNA linear GSS 21 JUN 2000
 DEFINITION mR182P RFLP sequences of mungbean, Vigna radiata
 genomic, DNA sequence.
 ACCESSION A2254240
 VERSION A2254240.1 GI:8602504
 KEYWORDS GSS.
 SOURCE Vigna radiata.
 ORGANISM Vigna radiata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Vigna.

REFERENCE 1 (bases 1 to 434)
 AUTHORS Denny, E., Daresch, D., Madge, J., Cooper, A., Larson, K.,
 Menancio-Hautea, D., Kumar, L. and Young, N. D.
 TITLE RFLP sequences of mungbean, Vigna radiata
 JOURNAL Unpublished (2000)
 COMMENT Contact: Young Nevin D
 Department of Plant Pathology
 University of Minnesota
 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA
 Tel: 612 625 2225
 Fax: 612 625 9728
 Email: nevin@tc.umn.edu

Sequence of mapped RFLP marker mR182 on linkage group 1 of
 mungbean; linkage group 11 of cowpea. For more information, see
 Beangenes at:
<http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=beangenes&cl=app=locus>. Please see authorities for mapping/naming: Fatokun,
 C. A., Daresch, D., Young, N. D. (1993) RFLP linkage map for cowpea
 (Vigna unguiculata (L.) Walp.) in: Genetic Maps 1992, S. J.
 O'Brien, ed. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp.
 6.256-6.258.
 Young, N. D. (1993) RFLP linkage map for mungbean (Vigna radiata (L.)
 Wilczek) in: Genetic Maps, 1992, S. J. O'Brien, ed. Cold Spring
 Harbor Press, Cold Spring Harbor, NY, pp. 6.259-6.260
 Insert length: 1300 Std Error: 0.00
 Seq primer: M18
 Class: RFLP probe.

FEATURES
 source
 Location/Qualifiers
 1..434
 /organism="Vigna radiata"
 /cultivar="Textsprout"
 /db_xref="taxon:157791"
 /clone_lib="RFLP sequences of mungbean, Vigna radiata"
 /tissue_type="Hypocotyl and roots"
 /dev_stage="Sprouts"
 /note="Vector: pUC 18; Site 1. Pst I, DNA was digested
 with Pst I, size separated by sucrose gradient
 centrifugation and the fraction between 500-3000 base
 pairs ligated into the vector using standard method."
 103 a 115 c 101 g 101 r 14 others

Query Match 1.2%; Score 36; DB 17; Length 434;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATTCACCTAGTAAACGGCGGCGGAGTGTGTGTGGAAT 36
 |||
 DB 18 CGGATTCACCTAGTAAACGGCGGCGGAGTGTGTGTGGAAT 53

RESULT 8
 LOCUS AL449722 451 bp mRNA linear EST 15 NOV 2000
 DEFINITION AL449722 Homo sapiens fetal brain (Stavrides GS) Homo sapiens cDNA,
 mRNA sequence.
 ACCESSION AL449722
 VERSION AL449722.1 GI:11191356
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Chordata; Cladacea; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 451)
 AUTHORS Stavrides, G. S., Huckle, E. J. and Deloukas, P.
 TITLE Unpublished (2000)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Stavrides GS
 The Sanger Centre
 Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: huckle@hinxton.sanger.ac.uk
 Sanger Centre name: scd1019.17.2fp.

FEATURES
 source
 Location/Qualifiers
 1..451
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="20"
 /clone_lib="Homo sapiens fetal brain (Stavrides GS)"
 /tissue_type="Brain"
 /dev_stage="fetal"
 /note="cDNA fragment isolated using a cDNA end rescue
 technique"

BASE COUNT 63 a 147 c 135 g 108 t
 ORIGIN

Query Match 1.2%; Score 46; DB 9; Length 451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATTCACCTAGTAAACGGCGGCGGAGTGTGTGTGGAAT 36
 |||
 DB 12 CGGATTCACCTAGTAAACGGCGGCGGAGTGTGTGTGGAAT 47

RESULT 9
 LOCUS BE414749 452 bp mRNA linear EST 24 JUL 2000
 DEFINITION MW1002 A06P900430 ITCF MW1 wheat Root Library Triticum aestivum
 cDNA clone MW1002.A06, mRNA sequence.
 ACCESSION BE414749
 VERSION BE414749.1 GI:9412651
 KEYWORDS EST.
 SOURCE Bread wheat.
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
 : Triticeae; Triticum.
 1 (bases 1 to 452)

AUTHORS Anderson, O. A., Appels, R., Bailey, P., Hake, J., Close, T., Cloutier
 S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
 Herrmann, P. G., Holton, T., Jacquemin, J. M., Jia, J., Joudrier, P.,
 Langridge, P., Lazo, G. R., Lin, J. J., McGuire, P., Ogilhard, Y.,
 Perichon, N., Quilset, C., Schuch, W., Solivar, G., Sharif, M.,
 Sorrells, M., Warburton, M. and Wenzel, G.
 TITLE International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)

JOURNAL Applied Biotechnology Center, CIMMYT
 COMMENT Apdo. Postal 6-641, 06600 Mexico DF MEXICO
 Tel: 52-5-7269091 ext 1381
 Fax: 52-5-726758/59
 Email: mwarburton@cnet.com
 International Triticeae EST Cooperative (ITEC)

Dept of Cellular Biology
 Institute of Basic Medical Sciences/Peking Union Medical College &
 Chinese Academy of Medical Sciences (PUMC & CAMS)
 5 Dong Dan San Tiao, Beijing, 100005 P R China
 Tel: 8601-65296459
 Fax: 8610-65240529
 Email: xuosp@cmi.imicams.ac.cn
 Seq primer: M13 Forward and Reverse Primers

FEATURES
 source
 location/Qualifiers

1. 503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal liver (S.Xue)"
 /tissue_type="liver"
 /dev_stage="fetal"
 /lab_host="E.coli DH5a"
 /note="Vector: pBluescript SK"

BASE COUNT 128 a 124 c 118 g 133 t

ORIGIN

Query Match 1.28; Score 36; DB 9; Length 503;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGGTGAAT 36
 |||||
 DB 24 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGGTGAAT 59

RESULT 14

W91680 555 bp mRNA linear EST 09-JUL-1996
 LOCUS MTA.C02.079.A MTA adult mouse thymus library Mus musculus cDNA
 DEFINITION clone MTA.C02.079 5' end similar to Human EST, mRNA sequence.

ACCESSION W91680

VERSION W91680

KEYWORDS EST.

SOURCE W91680 1 GI:1408034

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 555)

Authors Nguyen,C., Rocha,D., Granjeaud,S., Bernard,K., Naquet,P. and Jordan

,B.K.

Title Gene expression in different cell types of the mouse thymus

Journal Unpublished (1996)

Contact: Jordan BR

Genome Structure and Immune Functions

Centre d'Immunologie INSERM/CNRS

Case 906, 13288 MAPSEILLE Cedex 9, FRANCE

Tel: 330491269430

Fax: 330491269430

Email: jordan@cmi.univ-mrs.fr

This sequence was determined at Genome Express, Grenoble, France

Seq primer: T7

FEATURES

source

location/Qualifiers

1. 555

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="MTA.C02.079"

/clone_lib="MTA adult mouse thymus library"

/lab_host="MC1061 pi"

/note="Vector: pCDNA1; Site_1: NotI; Site_2: EcoRI; The

cDNA library was constructed from poly(A)+ RNA of an adult

mouse thymus by oligo dT primed reverse transcription.

cDNA was selected on gel for size above 800 nucleotides

after second strand synthesis then directionally cloned

into the pCDNA1 vector (NotI on polyA side, EcoRI on the

5'side)."

BASE COUNT 149 a 112 c 103 g 189 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 36; DB 14; Length 555;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGGTGAAT 36
 |||||
 DB 15 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGGTGAAT 50

RESULT 14

BM888419 612 bp mRNA linear EST 08 MAR 2002
 LOCUS TW077 Human Trabecular Meshwork cDNA library Homo sapiens cDNA
 DEFINITION clone 104138 5', mRNA sequence.

ACCESSION BM888419

VERSION BM888419.1 GI:19272163

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 612)

Authors Wirtz,M.K., Samples,I.P., Xu,H., Severson,T. and Acott,T.S.

Title Expression Profile and Genome Location of cDNA clones from an

Infant Human Trabecular Meshwork Library

Journal Unpublished (2002)

Contact: Wirtz MK

Glaucoma Genetics Lab

Oregon Health Sciences University

3375 S.W. Terwilliger Blvd., Portland, OR 97201-4147, USA

Tel: 503-494-4698

Fax: 503-494-6875

Email: wirtz@ohsu.edu

Seq primer: T7 Reverse

High quality sequence stop: 350.

FEATURES

source

location/Qualifiers

1. 612

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="104138"

/clone_lib="Human Trabecular Meshwork cDNA library"

/cell_type="trabecular meshwork"

/dev_stage="2 week to 2 year old infants"

/lab_host="Tc1010"

/note="Vector: pCDNA3; Site_1: EcoRI; Site_2: EcoRI; Human

cDNA library made from mRNA isolated from Trabecular

meshwork cells established from eyes of 6 individuals,

ages 2 weeks to 2 years. Cells were harvested at passages

3 through 6. Invitrogen made a unidirectional cDNA library

from the mRNA from the frozen cells using a pCDNA3 vector

and Tc1010 host cells."

BASE COUNT 158 a 127 c 145 g 142 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 36; DB 14; Length 612;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGGTGAAT 36
 |||||
 DB 9 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGGTGAAT 44

RESULT 15

BM887889 633 bp mRNA linear EST 08 MAR 2002
 LOCUS TM105 Human Trabecular Meshwork cDNA library Homo sapiens cDNA
 DEFINITION mRNA sequence.

ACCESSION BM887889

VERSION BM887889.1 GI:19271633

KEYWORDS EST.

```

SOURCE      human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 633)
AUTHORS      Wirtz, M. K., Samples, J. R., Xu, H., Severson, T. and Acott, T. S.
TITLE        Expression Profile and Genome Location of cDNA Clones from an
              Infant Human Trabecular Meshwork Library
JOURNAL      Unpublished (2002)
COMMENT      Contact: Wirtz MK
              Glaucoma Genetics Lab
              Oregon Health Sciences University
              3375 S.W. Terwilliger Blvd., Portland, OR 97201-4197, USA
              Tel: 503-494-4698
              Fax: 503-494-6875
              Email: wirtzm@ohsu.edu
              Seq primer: T7 Reverse.
FEATURES     Location/Qualifiers
              1..633
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="Human Trabecular Meshwork cDNA library"
               /tissue_type="eye"
               /cell_type="trabecular meshwork"
               /dev_stage="2 week to 2 year old infants"
               /lab_host="TOP10F"
               /note="vector: pcDNA3; Site_1: EcoRI; Site_2: EcoRI; Human
              cDNA library made from mRNA isolated from trabecular
              meshwork cells established from eyes from 6 individuals,
              ages 2 weeks to 2 years. Cells were harvested at passages
              3 through 6. Invitrogen made a unidirectional cDNA library
              from the mRNA from the frozen cells using a pcDNA3 vector
              and TOP10F host cells."
BASE COUNT   161 a 125 c 165 g 179 t      3 others
ORIGIN

Query Match      1.2%; Score 36; DB 14; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CGGATCCACTAGTAAAGGCGCGCTAGTGTGGTGGAAAT 36
        |||||||
Db      39 CGATCCACTAGTAAAGGCGCGCGCTAGTGTGGTGGAAAT 74

Search completed: July 16, 2003, 05:27:25
Job time : 4170 secs

```




```

Db 2341 AAACCAAGTACATATGGTTTATGTAAGCAATTAATTAATATACAGTGTATTCATCAATCA 2400
QY 2401 ATATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2450
Db 2401 ATGTTATATGAATTTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2460
QY 2461 GTTATATATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
Db 2461 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2520
QY 2521 TTAATTTGTTTACTATATATATATATATATATATATATATATATATATATATATAT 2580
Db 2521 TTAATTTGTTTACTATATATATATATATATATATATATATATATATATATATATAT 2580
QY 2581 GATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2640
Db 2581 GATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2640
QY 2641 AATCAGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
Db 2641 AATCAGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
QY 2701 AATCAGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
Db 2701 AATCAGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
QY 2761 ATGTCACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2820
Db 2761 ATGTCACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2820
QY 2821 CATCATATAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
Db 2821 CATCATATAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
QY 2881 CAGAGCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2940
Db 2881 CAGAGCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2940
QY 2941 TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Db 2941 TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
QY 3001 CATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
Db 3001 CATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
QY 3061 CCATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3111
Db 3061 CCATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3111

```

```

RESULT 2
US-10-201-386-6
; Sequence 6, Application US/10201386
; Publication No. US20030091567A1
; GENERAL INFORMATION:
; APPLICANT: Aitalev, Kari
; APPLICANT: Joukov, Vladimir
; TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR C (VEGF-C) PROTEIN
; TITLE OF INVENTION: AND GENES, MUTANTS THEREOF, AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/10201386
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US/05/544 376
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 05/455,700
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: PCT/US98/01973
; PRIOR FILING DATE: 1998-02-02
; PRIOR APPLICATION NUMBER: 06/755,430
; PRIOR FILING DATE: 1997-02-05
; PRIOR APPLICATION NUMBER: PCT/FR96/00427

```

```

; PRIOR FILING DATE: 1996-08-01
; PRIOR APPLICATION NUMBER: 08/671,573
; PRIOR FILING DATE: 1996-06-28
; PRIOR APPLICATION NUMBER: 08/601,132
; PRIOR FILING DATE: 1996-02-14
; PRIOR APPLICATION NUMBER: 08/585,895
; PRIOR FILING DATE: 1996-01-12
; PRIOR APPLICATION NUMBER: 08/510,133
; PRIOR FILING DATE: 1995-08-01
; PRIOR APPLICATION NUMBER: 08/440,011
; PRIOR FILING DATE: 1994-11-14
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 5
; LENGTH: 219
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: vector and
; OTHER INFORMATION: human VEGF-C cDNA
US-10-201-386-6

```

```

Query Match 1.28; Score 36; DB 9; Length 219;
Best Local Similarity 100.0%; Pred. No. 1 be-07;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CGGATCCACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 36
Db 35 CGGATCCACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 70

```

```

RESULT 3
US-09-933-797-116
; Sequence 116, Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikos et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/933,797
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US/05/482,933
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/085,483
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 116
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(380)
; OTHER INFORMATION: n = A, I, C or G
US-09-933-797-116

```

```

Query Match 1.28; Score 36; DB 9; Length 380;
Best Local Similarity 100.0%; Pred. No. 1 be-07;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CGGATCCACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 36
Db 35 CGGATCCACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 70

```

```

RESULT 4
US-09-933-797-89
; Sequence 89, Application US/09933797
; Patent No. US20020155119A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT FILING DATE: 2001-08-22
; PRIOR FILING DATE: 2001-08-22
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US/09/482,933
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/045,484
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 89
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Murine
; NAME/KEY: misc.feature
; LOCATION: (1)..(575)
; OTHER INFORMATION: n = A,T,C or G
US-09-933-797-89

Query Match      1.2%   Score 36;   DB 9;   Length 575;
Best local Similarity 100.0%;   Pred. No. 1.7e 07;
Matches 46;   Conservative 0;   Mismatches 0;   Indels 0;

QY 1 CGGATCCATAGTAAAGGGGGGGGGGCGATGCTGGTGAAT 36
      |||||||
DB 40 CGGATCCATAGTAAAGGGGGGGGGGCGATGCTGGTGAAT 65

RESULT 5
US-09-443-218-1
; Sequence 1, Application US/09443218
; Publication No. US20030094610A1
; GENERAL INFORMATION:
; APPLICANT: Lowenthal, John W
; APPLICANT: Johnson, Michael A
; APPLICANT: O'Neil, Terri E.
; APPLICANT: York, Jennifer J.
; TITLE OF INVENTION: No US20030094610A1-1 Uses of Avian Interferon Gamma
; FILE REFERENCE: 48-95C
; CURRENT FILING DATE: 1999-11-19
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT AU95/00114
; PRIOR FILING DATE: 1996-03-05
; PRIOR APPLICATION NUMBER: AU PM1542/95
; PRIOR FILING DATE: 1995-03-06
; PRIOR APPLICATION NUMBER: US 09/272,042
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Gallus sp. (chicken)
; NAME/KEY: CDS
; LOCATION: (14)..(625)
; NAME/KEY: mat_prime
; LOCATION: (191)..(625)
; NAME/KEY: 3'UTR
; LOCATION: (626)..(1079)
; NAME/KEY: 5'UTR
; LOCATION: (1)..(143)
US-09-443-218-1

Query Match      1.1%   Score 35;   DB 9;   Length 1079;

```

```

Best local Similarity 100.0%;   Pred. No. 5.9e-07;
Matches 35;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY 2 GGATCCATAGTAAAGGGGGGGGGGCGATGCTGGTGAAT 36
      |||||||
DB 1 GGATCCATAGTAAAGGGGGGGGGGCGATGCTGGTGAAT 35

RESULT 6
US-09-933-797-112
; Sequence 112, Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT FILING DATE: 2001-08-22
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,383
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Murine
; NAME/KEY: misc.feature
; LOCATION: (1)..(331)
; OTHER INFORMATION: n = A,T,C or G
US-09-933-797-112

Query Match      1.0%   Score 42;   DB 9;   Length 331;
Best local Similarity 100.0%;   Pred. No. 1.8e-05;
Matches 32;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY 1 CGGATCCATAGTAAAGGGGGGGGGGCGATGCTGGTGG 32
      |||||||
DB 50 CGGATCCATAGTAAAGGGGGGGGGGCGATGCTGGTGG 81

RESULT 7
US-09-796-6573-C
; Sequence 6573, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Mangan, John
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077 001200
; CURRENT FILING DATE: 2001-03-01
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,404
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04

```


PRIOR APPLICATION NUMBER: 60/222,416
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6573
LENGTH: 101
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (101)
OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6573

Query Match 0.9%; Score 29; DB 9; Length 101;
Best Local Similarity 100.0%; Pred No. 0.00053;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTACGCGCCGACGTGTG 29
|||||
Db 51 CGGATCCACTAGTACGCGCCGACGTGTG 29

RESULT 8

US-10-040-862-6573/C
Sequence 6573, Application US/10040862
Publication No. US2003007896A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
FILE REFERENCE: 014058-01352605
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US/10/040,862
PRIOR FILING DATE: 2000-04-01
PRIOR APPLICATION NUMBER: US/60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US/60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US/60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US/60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US/60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US/60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US/60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US/60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US/60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US/60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US/60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US/60/223,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6573
LENGTH: 101
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (101)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-6573

Query Match 0.9%; Score 29; DB 9; Length 101;
Best Local Similarity 100.0%; Pred No. 0.00053;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTACGCGCCGACGTGTG 29
|||||
Db 51 CGGATCCACTAGTACGCGCCGACGTGTG 29

RESULT 9

US-09-834-975-712
Sequence 712, Application US/09834975
Patent No. US20020110415A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Bolt, Andrew
APPLICANT: Van Hufel, Christophe
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-016B
CURRENT APPLICATION NUMBER: US/09/834,975
CURRENT FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/197,538
PRIOR FILING DATE: 2000-04-14
NUMRFP OF SEQ ID NOS: 1046
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 712
LENGTH: 101
TYPE: DNA
ORGANISM: Homo sapiens
US-09-834-975-712

Query Match 0.9%; Score 29; DB 19; Length 101;
Best Local Similarity 100.0%; Pred No. 0.00053;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTACGCGCCGACGTGTG 29
|||||
Eb 1 CGGATCCACTAGTACGCGCCGACGTGTG 29

RESULT 10

US-09-400-530A-11
Sequence 11, Application US/09900530A
Patent No. US20020128438A1
GENERAL INFORMATION:
APPLICANT: Seol, Dae-Wu
APPLICANT: Billari, Timothy R.
TITLE OF INVENTION: DNA Cassette for the Production of
Secretable Recombinant Trimeric Trail Proteins, Tetraacycline
Resistant Recombinant Trimeric Adeno-Associated Virus Vectors, Their
Use in Gene Therapy
FILE REFERENCE: 5006-1-002
CURRENT APPLICATION NUMBER: US/09/400,530A
CURRENT FILING DATE: 2002-02-19
PRIOR APPLICATION NUMBER: KR 2000-38441
PRIOR FILING DATE: 2000-07-06
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11

; LENGTH: 106
; TYPE: DNA
; ORGANISM: cytomegalovirus
US-09-900-530A-11

Query Match 0 98; Score 29; DR 10; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.00053;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGATCCACTAGTAACGGCGCCAGTGTG 29

DB 1B CGGATCCACTAGTAACGGCGCCAGTGTG 46

RESULT 11

US-09-864-864-67

; Sequence 67, Application US/098484

; Patent No. US20020102679A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Dillon, David C.

; APPLICANT: Secrist, Heather

; APPLICANT: Lodes, Michael J.

; APPLICANT: Algate, Paul A.

; APPLICANT: Fling, Steve P.

; APPLICANT: Mannion, Jane

; APPLICANT: Benson, Darin R.

; APPLICANT: Carter, Derrick

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.523

; CURRENT APPLICATION NUMBER: US/09/864,864

; CURRENT FILING DATE: 2001-05-23

; NUMBER OF SEQ ID NOS: 341

; SOFTWARE: Corixa Invention Disclosure Database

; SEQ ID NO 67

; LENGTH: 122

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(122)

; OTHER INFORMATION: n=A,T,C or G

US-09-864-864-67

Query Match

Best Local Similarity 100.0%; Pred. No. 0.00054;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGATCCACTAGTAACGGCGCCAGTGTG 29

DB 40 CGGATCCACTAGTAACGGCGCCAGTGTG 58

RESULT 12

US-09-796-692-2945/c

; Sequence 2945, Application US/09796692

; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; FILE REFERENCE: 2077-091200

; CURRENT APPLICATION NUMBER: US/09/796,692

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,403
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 4597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2945
; LENGTH: 138
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-2945

Query Match

Best Local Similarity 100.0%; Pred. No. 0.00054;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGATCCACTAGTAACGGCGCCAGTGTG 29

DB 91 CGGATCCACTAGTAACGGCGCCAGTGTG 63

RESULT 13

US-09-796-692-3006/c

; Sequence 3006, Application US/09796692

; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER

; FILE REFERENCE: 2077-091200

; CURRENT APPLICATION NUMBER: 60/200,545

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/200,904

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: 60/223,378



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 20:19:01 ; Search time 147 seconds
(without alignments)
10861.954 Million cell updates/sec

Title: US-10-053-192-2

Perfect score: 3111

Sequence: 1 cggatccactagtaacggcc.....aaaaaaaaaaaaaaaaaaaaa 3111

Scoring table: Gapped_Nuc

Gapop 60 0, Gapext 50.0

Search-hits: 2185234 seqs, 112599359 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N_geneseq_101062.2

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3111	100.0	3111	21 AAA53888	Beta, Beta-carotene
2	36	1.2	219	19 AAV48407	Human Flt4 ligand
3	36	1.2	1288	20 AAX25511	Human herpes simpl
4	36	1.2	198	19 AAV10456	Wheat low activity
5	35	1.1	1079	17 AAV38068	Chicken gamma inte
6	34	1.1	3136	16 AAV24264	Nucleotide sequen
7	34	1.1	3130	21 AAA12670	cDNA encoding huma
8	33	1.1	871	21 AA259142	Metastasis-associ
9	30	1.0	275	22 AAH22658	PPARGamma responsi

c	10	40	1.0	455	23 AAH22659	PPARGamma responsi
c	11	39	1.0	456	22 AAH22658	Human polynucleoti
c	12	39	1.0	1889	20 AAV69433	Human p53 cDNA #
c	13	29	0.9	48	21 AAA30408	M. tuberculosis Cu
c	14	29	0.9	49	24 ARA97964	IAF shuttle vector
c	15	29	0.9	44	19 AA299695	Vector back-end ol
c	16	29	0.9	44	16 AA084646	Vector back-end ol
c	17	29	0.9	151	22 AA269711	Human Cancer agent
c	18	29	0.9	103	22 AAH82475	Human ovarian tumo
c	19	29	0.9	114	24 ARL22963	Kat metastatic tum
c	20	29	0.9	122	24 ARL09530	Human ovarian tumo
c	21	29	0.9	143	22 AA204332	Human IREN 5' intr
c	22	29	0.9	148	24 AA056314	Human colon cancer
c	23	29	0.9	174	23 AA558839	cDNA #1515 encodin
c	24	29	0.9	185	22 AAH83439	Human ovarian tumo
c	25	29	0.9	195	22 AAH83432	Human ovarian tumo
c	26	29	0.9	196	24 ARL56000	Polynucleotide tes
c	27	29	0.9	198	22 AAH83749	Human ovarian tumo
c	28	29	0.9	201	22 AAH23830	Human ovarian tumo
c	29	29	0.9	207	22 AAH83487	Human ovarian tumo
c	30	29	0.9	210	23 AA557531	cDNA #207 encoding
c	31	29	0.9	211	22 AAH83472	Human ovarian tumo
c	32	29	0.9	212	22 AAH83402	Human ovarian tumo
c	33	29	0.9	217	21 AA397340	Human UHR22 5' cDN
c	34	29	0.9	217	22 AAH82572	Human ovarian tumo
c	35	29	0.9	220	22 AAH83656	Human ovarian tumo
c	36	29	0.9	223	24 ARL5732	Plasmid pCR2.1 Bar
c	37	29	0.9	223	22 AAH83148	Human ovarian tumo
c	38	29	0.9	226	22 AAH82043	Human ovarian tumo
c	39	29	0.9	227	22 AAH83003	Human ovarian tumo
c	40	29	0.9	229	24 ARL39244	cDNA encoding lung
c	41	29	0.9	232	22 AAH82412	Human ovarian tumo
c	42	29	0.9	233	23 AA577541	cDNA #217 encoding
c	43	29	0.9	234	22 AAH83590	Human ovarian tumo
c	44	29	0.9	235	22 AAH83645	Human ovarian tumo
c	45	29	0.9	235	23 AA557547	cDNA #223 encoding

ALIGNMENTS

RESULT 1

AAA53888

ID AAA53888 standard, DNA, 3111 BP.

XX

AC AAA53888;

XX

DF 03-JAN-2001 (first entry)

XX

RF Beta, Beta-carotene-15,15-dioxygenase coding sequence.

XX Beta, beta-carotene-15,15-dioxygenase, vitamin A, beta-carotene;

KW Trans-formate, fruit, vegetable, developmental disorder;

KW epithelial, fibroblast, antibody, detection, quantification;

KW treatment; therapy; ds.

XX Gallus gallus.

OS

XX Key Location/Qualifiers

XX CDS 133..1713

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PI Bachmann H, Bruenger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;
 E1 Wyss M.
 XX WPI: 2000 551036/51.
 DR p-PSDB: AAY97311.
 XX
 XX beta,beta-carotene 15,15'-dioxygenase protein, nucleic acids and
 PT antibodies, useful for production of vitamin A from carotene and gene
 PT therapy of ophthalmological disorders
 XX
 PS Claim 7: Page 14-15; 37pp; English.
 XX
 CC Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic
 CC conversion of beta-carotene to vitamin A and for raising specific
 CC antibodies. Nucleotides encoding all or part of bCOD are useful as
 CC primers or probes for specific amplification and/or detection of
 CC the gene that encodes bCOD, for isolation of related sequences in
 CC other organisms, for determining bCOD levels in humans (to identify
 CC subjects requiring vitamin A supplementation) and for detecting
 CC mutations in the bCOD gene. The nucleotide encoding bCOD may also
 CC be used to transform cells, particularly plant cells, to increase
 CC their vitamin A contents (especially in fruits and vegetables) and
 CC in gene therapy of subjects who have mutated or deleted forms of the
 CC bCOD gene and thus low vitamin A levels and susceptibility to
 CC developmental or ophthalmological disorders. Antibodies raised
 CC against bCOD are used for detection/quantification of bCOD in
 CC immunoassays.
 XX
 SQ Sequence 3111 BP; 1005 A, 620 C, 622 G, 864 T, 0 other.

QY 601 AAGGTACATACAGCAAAATATATAGTGTAAATTTGGCAATTTCTTCACTACTATATAG 660
 DB 601 AAGGTACATACAGCAAAATATATAGTGTAAATTTGGCAATTTCTTCACTACTATATAG 660
 QY 661 AGTGCTGCAATATTCCTCAACATGGTACTTCAATTTGTTGATAGAGAGAGAAATAATAT 720
 DB 661 AGTGCTGCAATATTCCTCAACATGGTACTTCAATTTGTTGATAGAGAGAGAAATAATAT 720
 QY 721 GTTCTCTTTAAGATCCCTTCTCTGATACAGAGAAAAAGAAAGCAAAATTTGTTTAAA 780
 DB 721 GTTCTCTTTAAGATCCCTTCTCTGATACAGAGAAAAAGAAAGCAAAATTTGTTTAAA 780
 QY 781 CACTGCAACTACTAGTCTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 DB 781 CACTGCAACTACTAGTCTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 QY 841 AGCTTTGSAATACAGAGAAAAATATATGTTGATAGAGAGAGAAATAATATATATAT 900
 DB 841 AGCTTTGSAATACAGAGAAAAATATATGTTGATAGAGAGAGAAATAATATATATAT 900
 QY 901 GTTAAATTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 DB 901 GTTAAATTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 QY 961 AAGGAGGATAGAGAGTGGTTCACCTTTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 DB 961 AAGGAGGATAGAGAGTGGTTCACCTTTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1021 AAGTTTACACGATGCTTTGGTGTGTATATATATATATATATATATATATATATATAT 1080
 DB 1021 AAGTTTACACGATGCTTTGGTGTGTATATATATATATATATATATATATATATATAT 1080
 QY 1081 CAGCTGCTTTTTCATATGCTTGGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 DB 1081 CAGCTGCTTTTTCATATGCTTGGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 QY 1141 AAAAACTGCAAG 1200
 DB 1141 AAAAACTGCAAG 1200
 QY 1201 GGTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 DB 1201 GGTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 QY 1261 CTTCACACTTCCGCAACTGCTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 DB 1261 CTTCACACTTCCGCAACTGCTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1321 ATATTATGTAAG 1380
 DB 1321 ATATTATGTAAG 1380
 QY 1381 AAGTATGCTATGCAACAG 1440
 DB 1381 AAGTATGCTATGCAACAG 1440
 QY 1441 AATGTCACAAAG 1500
 DB 1441 AATGTCACAAAG 1500
 QY 1501 ATCTTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 DB 1501 ATCTTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 QY 1561 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 DB 1561 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 QY 1621 GAATTGGGCGGAG 1680
 DB 1621 GAATTGGGCGGAG 1680

XX Gould-Rothberg BF;
 XX WPI; 2001-398172/42.
 XX Identifying ligands for peroxisome proliferator activated receptor
 PT gamma using differential expression of genes expressed in liver tissue
 PT
 XX
 XX Disclosure; Page 19; 84pp; English.
 XX The invention relates to methods of identifying ligands for peroxisome
 CC proliferator activated receptor gamma (PPARGamma) using differential gene
 CC expression. Also provided are PPARGamma responsive nucleotide sequences
 CC selected from HEPATO:1-18 and 44 or their complements. HEPATO are nucleic
 CC acid sequences differentially expressed in liver tissues of animals
 CC following exposure to N-(2-benzoylphenyl)-L-tyrosine, a synthetic
 CC PPARGamma ligand. The methods are useful for identifying a ligand for
 CC PPARGamma and a candidate therapeutic agent for a pathophysiology
 CC associated with a PPARGamma mediated metabolic pathway, such as
 CC noninsulin-dependent diabetes mellitus (NIDDM). The invention is also
 CC useful for diagnosing or determining susceptibility to a PPARGamma
 CC mediated pathophysiology e.g. NIDDM, liver failure, jaundice or ammonia
 CC toxicity. HEPATO sequences are useful for constructing probes for
 CC detecting HEPATO RNA sequences and for amplifying HEPATO sequences.
 CC HEPATO protein or its fragments is useful as an immunogen to generate
 CC antibodies that bind HEPATO. Sequences AAH22659-669 represent HEPATO
 CC nucleic acid sequences.
 XX Sequence 275 BP, 62 A, 74 C, 79 G, 57 T, 3 other;
 SQ
 Query Match: 1.0%; Score 36; 04 22; Length 376;
 Best Local Similarity 100.0%; Pred. No. 0 067;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 CGGATCCACTAGTAAAGGAGGCGGACGCGG 30
 Db 169 CGGATCCACTAGTAAAGGAGGCGGACGCGG 140
 RESULT 10
 AAH22659/C
 ID AAH22659 standard; DNA; 355 BP.
 XX
 AC AAH22659;
 XX
 XX 07-SEP-2001 (first entry)
 XX
 DE PPARGamma responsive novel HEPATO16 fragment.
 XX
 KW Peroxisome proliferator activated receptor gamma; PPARGamma; HEPATO;
 KW liver; N-(2-benzoylphenyl)-L-tyrosine, ligand; noninsulin-dependent
 KW gene therapy; noninsulin-dependent diabetes mellitus; jaundice;
 KW ammonia toxicity; ds.
 XX
 OS Rattus sp
 XX
 PN WO200144512-A2
 XX
 XX 21-JUN-2001.
 XX
 XX 15-DEC-2000; 2000WO-US34027.
 XX
 XX 16-DEC-1999; 99US-0171034.
 PR 06-FEB-2000; 2000US-0066547.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Gould-Rothberg BF;
 XX
 XX WPI; 2001-398172/42.
 XX
 XX Identifying ligands for peroxisome proliferator activated receptor

PT gamma using differential expression of genes expressed in liver tissue
 PT
 XX
 XX Disclosure; Page 19; 84pp; English.
 XX The invention relates to methods of identifying ligands for peroxisome
 CC proliferator activated receptor gamma (PPARGamma) using differential gene
 CC expression. Also provided are PPARGamma responsive nucleotide sequences
 CC selected from HEPATO:1-18 and 44 or their complements. HEPATO are nucleic
 CC acid sequences differentially expressed in liver tissues of animals
 CC following exposure to N-(2-benzoylphenyl)-L-tyrosine, a synthetic
 CC PPARGamma ligand. The methods are useful for identifying a ligand for
 CC PPARGamma and a candidate therapeutic agent for a pathophysiology
 CC associated with a PPARGamma mediated metabolic pathway, such as
 CC noninsulin-dependent diabetes mellitus (NIDDM). The invention is also
 CC useful for diagnosing or determining susceptibility to a PPARGamma
 CC mediated pathophysiology e.g. NIDDM, liver failure, jaundice or ammonia
 CC toxicity. HEPATO sequences are useful for constructing probes for
 CC detecting HEPATO RNA sequences and for amplifying HEPATO sequences.
 CC HEPATO protein or its fragments is useful as an immunogen to generate
 CC antibodies that bind HEPATO. Sequences AAH22659-669 represent HEPATO
 CC nucleic acid sequences.
 XX Sequence 355 BP, 71 A, 103 C, 110 G, 69 T, 2 other;
 SQ
 Query Match: 1.0%; Score 36; 04 22; Length 355;
 Best Local Similarity 100.0%; Pred. No. 0 065;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 CGGATCCACTAGTAAAGGAGGCGGACGCGG 30
 Db 169 CGGATCCACTAGTAAAGGAGGCGGACGCGG 220
 RESULT 11
 AAH83998
 ID AAH83998 standard; cDNA; 426 BP.
 XX
 AC AAH83998;
 XX
 XX 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 4058.
 XX
 KW Human; cytokine, cell proliferation, cell differentiation, gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 XX W020054835-A2.
 XX
 PD 07-SEP-2001.
 XX
 XX 26-FEB-2001; 2001WO-US04927.
 XX
 XX 28-FEB-2000; 2000US-015126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tanq YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-514449/56.
 DR P-PSDB; AA004067.
 XX
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 1; SEQ ID NO 4058; 1399pp + Sequence Listing; English.
 XX

CC The invention relates to human polypeptides (AA172941-AA193841) and
 CC the encoded proteins (AA000010-AA01910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulated
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 426 BP; 172 A; 60 C; 113 G; 65 T; 16 other;

Query Match 1.0%; Score 30; DB 22; Length 426;
 Best Local Similarity 100.0%; Pred. No. 0.003;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3082 CTCATATAAATAAAAAAAAAAAAAAAAAAAAA 3111
 Db 132 CTCATATAAATAAAAAAAAAAAAAAAAAAAAA 161

RESULT 12

AAV69433
 ID AAV69433 standard; cDNA; 1889 BP.

XX AAV69433;

DT 15-MAR-1999 (first entry)

DE Human ed4-6 cDNA #2.

XX Endothelial differentiation gene receptor homologues, human, HEDG;
 KW EDG-6; detection; diagnosis; drug screening; treatment; inflammation;
 KW disease; viral; bacterial; fungal infection; allergic response; injury;
 KW hereditary disease; lymphoma, carcinoma, lymphoid; neuronal; cascade;
 KW lymphocyte trafficking; leukocyte trafficking; signal transduction,
 KW rheumatoid synovium; autoimmune disorder; ss.

OS Homo sapiens.

PN W09854062-A1.

XX 26 Nov 1998.

PF 22-MAY-1998; 98WO-CA00487.

XX 22-MAY-1997; 97US-0861747.

XX (ALLX) ALLELIX BIOPHARMACEUTICALS INC.

PI Munroe DG, Vyas TB;

DR WPI; 1994-070147/06

PT New isolated human endothelial differentiation gene-6 receptor
 PT homologue - used to develop products for treating e.g. infections,
 PT allergic responses, trauma, hereditary diseases, lymphoma, carcinoma
 PT or autoimmune disorders

PS Disclosure; Fig 2; 59pp; English.

XX This sequence encodes a novel human endothelial differentiation gene
 CC (EDG)-6 receptor homologue designated HEDG. The encoded protein can be
 CC used for detection, diagnosis and drug screening. The protein can also
 CC be used to treat inflammation or diseases including viral, bacterial, or
 CC fungal infections, allergic responses, mechanical injury associated with
 CC trauma, hereditary diseases, lymphoma or carcinoma, or other conditions
 CC which activate the genes of lymphoid or neuronal tissues. It can also be

CC used for treating problems involving excessive lymphocyte and leukocyte
 CC trafficking. Inhibitors of HEDG are useful for controlling signal
 CC transduction and signalling cascades in cells of the rheumatoid synovium.
 CC The protein can also be used for treating autoimmune disorders such as
 CC myasthenia gravis.

XX Sequence 1889 BP; 300 A; 575 C; 537 G; 416 T; 1 other;

Query Match 1.0%; Score 30; DB 20; Length 1889;
 Best Local Similarity 100.0%; Pred. No. 0.052;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCATAGTAAATGGGCGGTCAGTGTGG 40
 Db 50 CGGATCCATAGTAAATGGGCGGTCAGTGTGG 79

RESULT 13

AAA40303/C

ID AAA40303 standard; DNA; 48 BP.

XX AAA40308;

DT 11-SEP-2000 (first entry)

DE M. tuberculosis Cu, Zn Superoxide dismutase, SOD, gene fragment.

XX Cu, Zn superoxide dismutase; antibody; tuberculosis; enzyme;

KW superoxide radical; macrophage; sodC; bacterium detection; ds.

OS Mycobacterium tuberculosis.

PN W0200029017-A1.

XX 25-MAY-2000.

XX 03-N-V-1999, 99W-MS0472.

PR 13-NOV-1998; 98US-0108309.

PA (YUNG-) YUNG SHIN PHARM IND CO LTD.

XX Lee FS, Wu CH;

XX WPI; 2000-38/619/33.

DR P-PSDB; AAY96264.

XX Novel monoclonal antibodies targeted to Mycobacterium tuberculosis
 PT superoxide dismutase, useful for detecting tuberculosis infection

XX Example 1; Page 14; 28pp; English.

XX The present sequence is a gene fragment of Mycobacterium tuberculosis
 CC Copper, Zinc Superoxide dismutase, Cu, ZnSOD, superoxide dismutase
 CC catalyses the conversion of superoxide radicals, which are mutagenic,
 CC into molecular oxygen and hydrogen peroxide. Cu, ZnSOD is thought to be
 CC important for bacterial survival in macrophages. Macrophages engulf
 CC bacteria, and set about destroying them by generating superoxide
 CC radicals. The bacterium counteracts by releasing SOD, to neutralise the
 CC superoxide radicals. The present sequence is added to the truncated
 CC Cu, ZnSOD coding sequence, (AAA40303) to produce a full-length version.
 CC The T of the stop codon of AAA40303 is mutated to an A, to produce a Lys
 CC codon. The present sequence was then added downstream of the Lys codon.
 CC Antibodies can be developed which bind specifically to the full length
 CC protein. The antibodies are useful for detecting tuberculosis
 CC infection in animals, and for detecting the presence of Mycobacterium
 CC tuberculosis. The antibodies are also used to isolate the Mycobacterium
 CC tuberculosis Cu, Zn SOD.

XX Sequence 48 BP; 10 A; 15 C; 13 G; 10 T; 0 other;

Query Match 0.9%; Score 29; DB 21; Length 48;
 Best Local Similarity 100.0%; Pred No. 0.2;

Matches 29, Conservative 0, Mismatches 0, Indels 0, Gaps 0.

QY 1 CGGATCCACCTAGTAACGGCGCCGACAGTGTG 29
 DB 40 CGGATCCACCTAGTAACGGCGCCGACAGTGTG 12

RESULT 14
 ABA97964
 ID ABA97964 standard; DNA: 49 BP.
 XX
 AC ABA97964:
 XX
 DT 25-APR-2002 (first entry)
 XX
 DE LAB shuttle vector multiple cloning site sequence SEQ ID NO 7.
 XX
 KW Lactic acid bacterium; shuttle vector; LAB: DNA vaccine;
 XX multiple cloning site; ds.
 XX
 OS Synthetic.
 XX
 PN JP2001340090-A.
 PD 11-DEC-2001.
 XX
 PF 09-MAR-2001; 2001JP-0067675.
 XX
 PP 26-MAY-2000; 2000TW-011025
 XX
 PA (ANAR-) ANARATA BIOTECH CORP LTD.
 XX
 DR WPI, 2002-135978/18.
 XX
 XX A lactic acid bacterium shuttle vector useful as a DNA vaccine carrier
 PT comprises a nucleic acid sequence encoding a protein related to
 PT replication of the LAB plasmid and a non-antibiotic-resistant selective
 PT marker -
 XX
 PS Example 5, Fig 4, 24pp; Japanese.
 XX
 CC The invention relates to a lactic acid bacterium (LAB) shuttle vector
 CC comprising:
 CC (a) LAB plasmid sequence containing the plus starting point of
 CC replication and a nucleic acid sequence encoding a protein related to
 CC replication of the LAB plasmid; and
 CC (b) a non-antibiotic-resistant selective marker and its promoter
 CC sequence. The lactic acid bacterium shuttle vector can be used as a DNA
 CC vaccine carrier, in selecting transformed cells, drugs and foods. The
 CC present sequence is that of a multiple cloning site of one of the LAB
 CC shuttle vectors of the invention.
 XX
 SQ Sequence 49 BP; 10 A; 14 G; 11 T; 0 other;

Query Match 0.9%; Score 29; DB 24; Length 49;
 Best Local Similarity 100.0%; Pred. No. 0.2;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACCTAGTAACGGCGCCGACAGTGTG 29
 DB 11 CGGATCCACCTAGTAACGGCGCCGACAGTGTG 39

RESULT 15
 AAQ89695/c
 ID AAQ89695 standard; DNA: 94 BP.
 XX
 AC AAQ89695;
 XX
 DT 07-NOV-1995 (first entry)
 XX
 DE Vector back-end oligonucleotide.
 XX

KW Universal cloning vector for expression of heterologous genes or
 XX back-end oligonucleotide; ss.
 OS Synthetic.
 XX
 PN WO9510620-A.
 XX
 PD 20-APR-1995.
 XX
 PF 14-OCT-1994; 94WO-US11/19.
 XX
 PP 15-OCT-1994; 94US-0136148.
 XX
 PA (MLCW) MALLINCKRODT VETERINARY INC.
 XX
 PI McMullen JR, Synenki RM, Zook CA;
 XX
 PP WPI; 1994-161805/21.
 XX
 PT Universal cloning vector for expression of heterologous genes or
 PT cDNA - allows high levels of expression without the need for
 PT sequence modification
 XX
 PS Claim 9; Fig.3; 40pp; English.
 XX
 CC A universal vector comprises a PL promoter, a ribosome binding site,
 CC the first 30 bp of the delta-7 pig somatostatin gene, a 21 bp front-
 CC end oligonucleotide (pref. the dimer comprising the complementary
 CC strands given in AAQ89695-94) containing multiple start and restriction
 CC enzyme sites, a back-end oligonucleotide (pref. the dimer
 CC comprising the strands given in AAQ89695-96) containing at least 1
 CC restriction enzyme site, a transcription terminator (pref. the dimer
 CC comprising the strands given in AAQ89695-98) and a drug resistance
 CC marker.
 XX
 SQ Sequence 94 BP; 20 A; 26 G; 20 C; 20 T; 0 other;

Query Match 0.9%; Score 29; DB 16; Length 94;
 Best Local Similarity 100.0%; Pred. No. 0.19;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACCTAGTAACGGCGCCGACAGTGTG 29
 DB 82 CGGATCCACCTAGTAACGGCGCCGACAGTGTG 54

Search completed: July 16, 2003, 00:14:45
 Job time : 647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 00:03:57, Search time 147 seconds
(without alignments)
6964 025 Million cell updates/sec

Title: US-10-053-192-2

Perfect score: 3111

Sequence: 1 cygatccactagtaacggcc

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size: 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A.COMR seq.*
- 2: /cgn2_6/ptodata/1/ina/5R.COMR seq.*
- 3: /cgn2_6/ptodata/1/ina/6A.COMR seq.*
- 4: /cgn2_6/ptodata/1/ina/6R.COMR seq.*
- 5: /cgn2_6/ptodata/1/ina/PTCUS.COMR seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DR	ID	Description
1	36	1.2	219	3	US-08-795-430-6	Sequence 5, Appli
2	36	1.2	219	4	US-08-510-133A-25	Sequence 25, Appli
3	36	1.2	219	4	US-08-585-895-25	Sequence 25, Appli
4	36	1.2	219	4	US-09-355-700-6	Sequence 6, Appli
5	36	1.2	219	4	US-08-601-132-25	Sequence 25, Appli
6	36	1.2	1982	2	US-08-900-148-1	Sequence 1, Appli
7	35	1.1	1079	3	US-08-765-381-1	Sequence 1, Appli
8	35	1.1	1302	1	US-08-525-697-1	Sequence 1, Appli
9	33	1.1	870	4	US-09-333-599-1	Sequence 1, Appli
10	30	1.1	1889	3	US-08-861-747-3	Sequence 3, Appli
11	29	0.9	115	2	US-08-454-577C-75	Sequence 75, Appli
12	29	0.9	115	2	US-08-340-426D-75	Sequence 75, Appli
13	29	0.9	115	2	US-08-450-673C-75	Sequence 75, Appli
14	29	0.9	115	5	US-08-17111A-75	Sequence 75, Appli
15	29	0.9	147	4	US-08-604-165-17	Sequence 17, Appli
16	29	0.9	147	4	US-08-734-054B-17	Sequence 17, Appli
17	29	0.9	151	4	US-09-359-361-7	Sequence 7, Appli
18	29	0.9	170	4	US-08-604-165-16	Sequence 16, Appli
19	29	0.9	170	4	US-08-734-054B-16	Sequence 16, Appli
20	29	0.9	371	4	US-09-404-879A-365	Sequence 365, App
21	29	0.9	392	4	US-09-404-879A-381	Sequence 381, App
22	29	0.9	434	4	US-09-404-879A-461	Sequence 461, App
23	29	0.9	396	4	US-09-404-879A-383	Sequence 383, App
24	29	0.9	404	1	US-08-594-031-154	Sequence 154, App
25	29	0.9	415	4	US-09-404-879A-365	Sequence 365, App
26	29	0.9	448	4	US-08-821-827C-15	Sequence 15, Appli
27	29	0.9	440	4	US-09-290-202B-15	Sequence 15, Appli

C 28	29	0.9	741	1	US-09-616-468A-2	Sequence 2, Appli
C 29	29	0.9	741	4	US-09-054-298-2	Sequence 2, Appli
C 30	29	0.9	741	4	US-08-818-635-2	Sequence 2, Appli
C 31	29	0.9	744	4	US-09-175-658B-24	Sequence 24, Appli
C 32	29	0.9	752	4	US-09-175-658B-23	Sequence 23, Appli
C 33	29	0.9	908	4	US-08-991-789A-46	Sequence 46, Appli
C 34	29	0.9	908	4	US-08-062-451-46	Sequence 46, Appli
C 35	29	0.9	908	4	US-09-594-456-46	Sequence 46, Appli
C 36	29	0.9	942	4	US-09-879-027-5	Sequence 5, Appli
C 37	29	0.9	1084	4	US-08-460-048-3	Sequence 3, Appli
C 38	29	0.9	1123	1	US-08-458-023B-3	Sequence 3, Appli
C 39	29	0.9	1186	4	US-09-378-898A-89	Sequence 89, Appli
C 40	29	0.9	1233	4	US-09-015-188-4	Sequence 4, Appli
C 41	29	0.9	1272	4	US-08-909-758-3	Sequence 3, Appli
C 42	29	0.9	1362	3	US-08-333-425-4	Sequence 4, Appli
C 43	29	0.9	1410	4	US-09-378-088A-81	Sequence 81, Appli
C 44	29	0.9	1590	4	US-08-908-738-4	Sequence 4, Appli
C 45	29	0.9	1654	4	US-09-216-909-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-795-430-6
Sequence 5, Affiliation US/08795430
Patent No. 6130071
GENERAL INFORMATION:
APPLICANT: Alltalo, Kari
TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)
TITLE OF INVENTION: Protein and Gene, Mutants Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08795430
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FI96/00427
FILING DATE: 01-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/471,573
FILING DATE: 28-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/401,122
FILING DATE: 14-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/585,895
FILING DATE: 12-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/510,133
FILING DATE: 01-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/340,011
FILING DATE: 14-NOV-1994
ALIGNMENT/ASMT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
PREFERENTIAL PRIORITY NUMBER: 244,672,3691
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300

OTHER INFORMATION: /product- "low affinity"ation
OTHER INFORMATION: transporter 1 (LCT1)"
US-08 900-148-1

Query Match 1.2% Score 36; DB 2; Length 1982;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 36
|||||
DB 18 CGGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 53

RESULT 7

US-08-765-481-1
Sequence 1, Application US/08765381
Patent No. 6083724

GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic
TITLE OF INVENTION: sequences encoding same
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:

ADDRESSEE: Scully Scott Murphy and Presser
STREET: 400 Garden City Plaza
CITY: Garden City, New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1 0, Version #1 25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765.381
FILING DATE: 19-DEC-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU P11542/95
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU96/00114
FILING DATE: 05-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: PRESSER, LEOPOLD
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-516-742-4343
TELEFAX: 1-516-742-4366

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1079 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Chicken (Gallus sp.)
CELL TYPE: T-cell
CELL LINE: CC8.1h
IMMEDIATE SOURCE:
LIBRARY: CC8.1h
CLONE: CHCN-gamma

FEATURE:

NAME/KEY: CDS
LOCATION: 134..625
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 191..625
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..133
FEATURE:
NAME/KEY: 3'UTR

LOCATION: 626..1079
US-08-765-381-1

Query Match 1.1% Score 35; DB 3; Length 1079;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 36
|||||
DB 1 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 35

RESULT 8

US-08-525-697-1

Sequence 1, Application US/08525697
Patent No. 5745764

GENERAL INFORMATION:
APPLICANT: Christgau, Stephan
APPLICANT: Andersen, Lene N
APPLICANT: Kauppinen, Sakari
APPLICANT: Helld-Hansen, Hans P
APPLICANT: Dalboeder, Henrik
TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5795764; No. 5795764disk of No. 5795764th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525.697
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/INDEX NUMBER: 4004.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-8123
TELEFAX: 212-878-9655

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 0486/93
FILING DATE: 30-APR-1993
CLASSIFICATION: 435

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1302 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 61..1192
US-08-525-697-1

Query Match 1.1% Score 35; DB 1; Length 1402;

Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 36
|||||
DB 1 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 35

RESULT 9

US-09-333-599-1

```
Sequence 1, Application US/09333599
Patent No. 6245898
GENERAL INFORMATION:
APPLICANT: Testa, Jacqueline E.
APPLICANT: Quigley, James P.
APPLICANT: Seandel, Marco
TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
FILE REFERENCE: SUNY
CURRENT APPLICATION NUMBER: US/09/333,599
CURRENT FILING DATE: 1999-06-15
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 870
TYPE: DNA
ORGANISM: Homo sapiens
US-09-333-599-1

Query Match      1.1%, Score 33, DB 4, Length 870,
Best Local Similarity 100.0%, Pred. No. 4 90-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATCCACTAGTAAAGGCGGAGTGTGTGTGAAT 36
    |||||
DB 1 ATCCACTAGTAAAGGCGGAGTGTGTGTGAAT 33

RESULT 10
US-08-861-747-3
Sequence 3, Application US/08861747
Patent No. 6020158
GENERAL INFORMATION:
APPLICANT: MUNROE, Donald G.
APPLICANT: VIAS, Tejal B.
TITLE OF INVENTION: A HUMAN EDG-6 PEPTIDE H-401
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikolov, Marmelstein, Murray & Gram LLP
STREET: 655 15th St., NW, Suite 330 - G Street Lobby
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/861,747
FILING DATE: 22-MAY-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jahus, Kristina M.
REGISTRATION NUMBER: 41,092
REFERENCE/POCKET NUMBER: PR074-7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-861-747-3

Query Match      1.0%, Score 30; DB 3; Length 1889;
Best Local Similarity 100.0%, Pred. No. 0.00069;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGCGGAGTGTGTGAAT 30
    |||||
DB 50 CGGATCCACTAGTAAAGGCGGAGTGTGTGAAT 79

Sequence 1, Application US/08454557C
Patent No. 5810670
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
APPLICANT: Wands, Jack R.
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,557C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/POCKET NUMBER: 0609,3840003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2500
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-454-557C-75

Query Match      0.9%, Score 29; DB 2; Length 115;
Best Local Similarity 100.0%, Pred. No. 0.003;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGCGGAGTGTGTGAAT 29
    |||||
DB 62 CGGATCCACTAGTAAAGGCGGAGTGTGTGAAT 34

Sequence 1, Application US/08340426D
Patent No. 5948634
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
APPLICANT: Wands, Jack R.
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/450,426D
 FILING DATE: 14-NOV-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, Steven R.
 REGISTRATION NUMBER: 36,203
 REFERENCE/DOCKET NUMBER: 0609,384,0002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 75:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 115 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: both
 US-08-450-426D-75

Query Match 0.99; Score 29; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGCGAGTGTG 29
 |||||
 DB 62 CGGATCCACTAGTAAAGCGGCGGCGAGTGTG 34

RESULT 13
 US-08-450-673C-75/c
 Sequence 75, Application US/08450673C
 Patent No. 5948888
 GENERAL INFORMATION:
 APPLICANT: de la Monte, Suzanne
 TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection
 TITLE OF INVENTION: of Alzheimer's Disease
 NUMBER OF SEQUENCES: 121
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/450,673C
 FILING DATE: 30-MAY-1995
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, Steven R.
 REGISTRATION NUMBER: 36,204
 REFERENCE/DOCKET NUMBER: 0609,384,0004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 75:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 115 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: both
 US-08-450-673C-75

Query Match 0.99; Score 29; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 0.003;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGCGAGTGTG 29
 |||||
 DB 62 CGGATCCACTAGTAAAGCGGCGGCGAGTGTG 34

RESULT 14
 PCT-US95-17111A-75/c
 Sequence 75, Application PC/TUS9517111A
 GENERAL INFORMATION:
 APPLICANT: de la Monte, Suzanne
 TITLE OF INVENTION: Neutral Thread Protein Gene Expression and
 TITLE OF INVENTION: Detection of Alzheimer's Disease
 NUMBER OF SEQUENCES: 121
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/17111A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/340,426
 FILING DATE: 14-NOV-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, Steven R.
 REGISTRATION NUMBER: 36,203
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 75:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 115 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: both
 PCT-US95-17111A-75

Query Match 0.99; Score 29; DB 5; Length 115;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGCGAGTGTG 29
 |||||
 DB 62 CGGATCCACTAGTAAAGCGGCGGCGAGTGTG 34

RESULT 15
 US-08-604-165-17
 Sequence 17, Application US/08604165
 Patent No. 6287557
 GENERAL INFORMATION:
 APPLICANT: Boursnell, Michael E.G.
 TITLE OF INVENTION: Viral Preparations, Vectors,
 TITLE OF INVENTION: Immunogens, and Vaccines
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,165
FILING DATE: 21-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9503395.7
FILING DATE: 21-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GR 9515557 8
FILING DATE: 28-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9503322.0
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Linda R. Judge
REGISTRATION NUMBER: 42,702
REFERENCE/DOCKET NUMBER: A-63282/WHD/LRJ
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 871-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 147 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
OS-08-604-165-17

[illegible]

Search completed: July 16, 2003, 05:30:02
Job time : 148 secs



Host Local Similarity 100 0%: Prod No 0:
Matches 3111: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

OY 1 CGGATCCATTAACGGCGCTGATGTGTGTAATCCATCTCTATAGTAACAGGAAG 60
Db 1 CGGATCCATTAACGGCGCGCCCACTGTGTGTAATCCATCTCTATAGTAACAGGAAG 60
OY 61 ACCTGTTTACCGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 61 ACCTGTTTACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
OY 121 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 121 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
OY 181 AAAGTGGAGTGGCAAGGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 181 AAAGTGGAGTGGCAAGGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
OY 241 CCAGGAGATGACACACATAGGGGACACTAATATATATATATATATATATATATAT 300
Db 241 CCAGGAGATGACACACATAGGGGACACTAATATATATATATATATATATATATAT 300
OY 301 CTGACAGCTTCACCTTAAATATGTAAGTATTAACAGAGAGAGAGAGAGAGAGAG 360
Db 301 CTGACAGCTTCACCTTAAATATGTAAGTATTAACAGAGAGAGAGAGAGAGAGAG 360
OY 361 GACACATACAACTGCAATATATATATATATATATATATATATATATATATATAT 420
Db 361 GACACATACAACTGCAATATATATATATATATATATATATATATATATATATAT 420
OY 421 GCTATATGCGATGCTATGCAAAACATATTTGCGAGAGAGAGAGAGAGAGAGAG 480
Db 421 GCTATATGCGATGCTATGCAAAACATATTTGCGAGAGAGAGAGAGAGAGAGAG 480
OY 481 ATTCTGAGTTCACGAGACACTGCTGATCAATATATATATATATATATATATATAT 540
Db 481 ATTCTGAGTTCACGAGACACTGCTGATCAATATATATATATATATATATATATAT 540
OY 541 GTTACGAGTGAAGTAACTTATATATATATATATATATATATATATATATATATAT 600
Db 541 GTTACGAGTGAAGTAACTTATATATATATATATATATATATATATATATATATAT 600
OY 601 AAGGTAGACTACAGCAATATATATATATATATATATATATATATATATATATAT 660
Db 601 AAGGTAGACTACAGCAATATATATATATATATATATATATATATATATATATAT 660
OY 661 AGTGTGGAATATTTTCAACATGAGTACTTCAATTTGATATATATATATATATAT 720
Db 661 AGTGTGGAATATTTTCAACATGAGTACTTCAATTTGATATATATATATATATAT 720
OY 721 GTTCTTTTATAGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 721 GTTCTTTTATAGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
OY 781 GACTTGAAGTATATATATATATATATATATATATATATATATATATATATATAT 840
Db 781 GACTTGAAGTATATATATATATATATATATATATATATATATATATATATATAT 840
OY 841 AGCTTTGGAATACAGAGAAATATATATATATATATATATATATATATATATATAT 900
Db 841 AGCTTTGGAATACAGAGAAATATATATATATATATATATATATATATATATATAT 900
OY 901 GTCAAGTGGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 901 GTCAAGTGGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
OY 961 AAGGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 AAGGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
OY 1021 AAGTTTTACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1021 AAGTTTTACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080

```

```

Db 1021 AAGTTTTACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
OY 1081 CAGGTGTTTTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1081 CAGGTGTTTTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
OY 1141 AAAAAGTGGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1141 AAAAAGTGGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
OY 1201 CGCTTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 1201 CGCTTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
OY 1261 CTTCGAAGTGGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1261 CTTCGAAGTGGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
OY 1321 ATATTATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 ATATTATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
OY 1381 AAGTATGCTATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 AAGTATGCTATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
OY 1441 AATGCTCAAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 AATGCTCAAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
OY 1501 ATCTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1501 ATCTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
OY 1561 GTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1561 GTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
OY 1621 GAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 1621 GAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
OY 1681 CCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1681 CCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
OY 1741 GAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Db 1741 GAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
OY 1801 AAGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db 1801 AAGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
OY 1861 AGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1861 AGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
OY 1921 AGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Db 1921 AGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
OY 1981 GGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
Db 1981 GGAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
OY 2041 TATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
Db 2041 TATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
OY 2101 GATGCTTTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
Db 2101 GATGCTTTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160

```

QY	2141	AAATGTCGGTAACTGCTAAATCTCAATCAATCAATCTCTGGTGGTGGATCTCAATCAATA	2250
Db	2141	AAATGTCGGTAACTGCTAAATCTCAATCAATCAATCTCTGGTGGTGGATCTCAATCAATA	2250
QY	2221	ACCAAAAGTTAGGTATTAATTCAGATGGCTTAATTTTAAATTCAGACCAAAACAAACA	2280
Db	2221	ACCAAAAGTTAGGTATTAATTCAGATGGCTTAATTTTAAATTCAGACCAAAACAAACA	2280
QY	2281	ACCAAAAGCAATCTGCTGCTTAATTAAGCAACCAACCAACCAACCAACCAACCAACCA	2340
Db	2281	ACCAAAAGCAATCTGCTGCTTAATTAAGCAACCAACCAACCAACCAACCAACCAACCA	2340
QY	2341	AAACCAAGATACAAATGCTTATATGATCAATTAATTAATTAATTAATTAATTAATTA	2400
Db	2341	AAACCAAGATACAAATGCTTATATGATCAATTAATTAATTAATTAATTAATTAATTA	2400
QY	2401	ATGTTATTAAGCAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2460
Db	2401	ATGTTATTAAGCAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2460
QY	2461	ATGTTATTAAGCAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2520
Db	2461	ATGTTATTAAGCAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2520
QY	2521	TTAATTTGTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2580
Db	2521	TTAATTTGTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2580
QY	2581	GATATATCAATCAATTAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2640
Db	2581	GATATATCAATCAATTAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2640
QY	2641	AAATGTCGGTAACTGCTAAATCTCAATCAATCAATCTCTGGTGGTGGATCTCAATCAATA	2700
Db	2641	AAATGTCGGTAACTGCTAAATCTCAATCAATCAATCTCTGGTGGTGGATCTCAATCAATA	2700
QY	2701	ATGTTATTAAGCAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2760
Db	2701	ATGTTATTAAGCAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2760
QY	2761	TTAATTTGTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2820
Db	2761	TTAATTTGTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2820
QY	2821	GATATATCAATCAATTAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2880
Db	2821	GATATATCAATCAATTAATTTGATTTAAATGAGTTTCTCTGAGATATCTGACATTAATTTT	2880
QY	2881	CAAGACGACCAATTAAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	2940
Db	2881	CAAGACGACCAATTAAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	2940
QY	2941	TCCTGCTCTTATGTTGATTTGATTTGAGCAATCACTGGCTAATTTAAATGTTAAAAAGT	3000
Db	2941	TCCTGCTCTTATGTTGATTTGATTTGAGCAATCACTGGCTAATTTAAATGTTAAAAAGT	3000
QY	3061	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3120
Db	3061	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3120
QY	3121	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3180
Db	3121	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3180
QY	3181	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3240
Db	3181	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3240
QY	3241	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3300
Db	3241	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3300
QY	3301	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3360
Db	3301	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3360
QY	3361	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3420
Db	3361	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3420
QY	3421	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3480
Db	3421	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3480
QY	3481	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3540
Db	3481	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3540
QY	3541	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3600
Db	3541	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3600
QY	3601	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3660
Db	3601	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3660
QY	3661	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3720
Db	3661	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3720
QY	3721	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3780
Db	3721	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3780
QY	3781	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3840
Db	3781	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3840
QY	3841	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3900
Db	3841	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3900
QY	3901	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3960
Db	3901	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	3960
QY	3961	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4020
Db	3961	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4020
QY	4021	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4080
Db	4021	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4080
QY	4081	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4140
Db	4081	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4140
QY	4141	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	4200
Db	4141	CAATTAATAAATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATA	

[illegible]

```
UY 541 GTATCACTGAGACTAACTTCATGCAAGAAATATGATCCAGACTCTTGACAGACTAGAT 609
DB 541 GCTACAGCTGAGACTAACTTCATGCAAGAAATATGATCCAGACTCTTGACAGACTAGAT 609
OY 601 AAGGTAGACTACAGCAAAATATAGTGTAACTTGGCAACTCTCCACCACTATGAC 660
DB 601 AAGGTAGACTACAGCAAAATATAGTGTAACTTGGCAACTCTCTCCACCACTATGAC 660
OY 661 AGCTGGAATATATTCATCATGAGTGGTACTTCAATTTGTATAGGAGCAACAAATAT 720
DB 661 AGCTGGAATATATTCATCATGAGTGGTACTTCAATTTGTATAGGAGCAACAAATAT 720
OY 721 GTTCCTTTAAGATCCCTTCCTCTGTACCAAGAAAGAAAGAAATCTTTTAA 780
DB 721 GTTCCTTTAAGATCCCTTCCTCTGTACCAAGAAAGAAAGAAATCTTTTAA 780
OY 781 CACCTGCAAGTATGATGCTGATCCCTTCCTCTGTACCAAGTACTACAC 840
DB 781 CACCTGCAAGTATGATGCTGATCCCTTCCTCTGTACCAAGTACTACAC 840
OY 841 AGCTTTGCAATCAGCAAAATATATATGTTGATAGAGGAGCATTAACTGATAT 900
DB 841 AGCTTTGCAATCAGCAAAATATATATGTTGATAGAGGAGCATTAACTGATAT 900
OY 901 GTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
DB 901 GTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
OY 961 AAGAGAGATAGAGCTGTTTACATTTTGTACCAAGAAAGAAAGAAATGATC 1020
DB 961 AAGAGAGATAGAGCTGTTTACATTTTGTACCAAGAAAGAAAGAAATGATC 1020
OY 1021 AAGTTTACACTGATGCTTGTGCTTATACCAATATATGCTTACGAGAGATG 1080
DB 1021 AAGTTTACACTGATGCTTGTGCTTATACCAATATATGCTTACGAGAGATG 1080
OY 1081 CACGTTGTTTGTATGCTTGTGCTTATACGAGAGATATGCTTACGAGAT 1140
DB 1081 CACGTTGTTTGTATGCTTGTGCTTATACGAGAGATATGCTTACGAGAT 1140
OY 1141 AAAAAAGTGCAGCAAGACTTGAAGTGAACAGCAAGTGAAGTGAAGTGAAG 1200
DB 1141 AAAAAAGTGCAGCAAGACTTGAAGTGAACAGCAAGTGAAGTGAAGTGAAG 1200
OY 1201 GCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
DB 1201 GCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
OY 1261 CTTGCAACTTCGCAACTGCTGTAAAGAAAGAAAGATGCGACATCTATGCTG 1320
DB 1261 CTTGCAACTTCGCAACTGCTGTAAAGAAAGAAAGATGCGACATCTATGCTG 1320
OY 1321 ATATTATGTAAGGATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 ATATTATGTAAGGATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
OY 1381 AAGTATGCTATGTAAGGATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 AAGTATGCTATGTAAGGATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
OY 1441 AATGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1500
DB 1441 AATGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1500
OY 1501 AATTTTCTTCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1560
DB 1501 AATTTTCTTCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1560
OY 1561 GTGCTGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1620
DB 1561 GTGCTGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1620
OY 1621 GAATTTGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1680
DB 1621 GAATTTGCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1680
OY 1681 CCACAGCAATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
DB 1681 CCACAGCAATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
OY 1741 GAGCAATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
DB 1741 GAGCAATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
OY 1801 AAGCAAGCAATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
DB 1801 AAGCAAGCAATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
OY 1861 AGCGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
DB 1861 AGCGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
OY 1921 AGGAGAGGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
DB 1921 AGGAGAGGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
OY 1981 GAGAGAGGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
DB 1981 GAGAGAGGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
OY 2041 TATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
DB 2041 TATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
OY 2101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
DB 2101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
OY 2161 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
DB 2161 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
OY 2221 AAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
DB 2221 AAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
OY 2281 AAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
DB 2281 AAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
OY 2341 AAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
DB 2341 AAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
OY 2401 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
DB 2401 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
OY 2461 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
DB 2461 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
OY 2521 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
DB 2521 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
OY 2581 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
DB 2581 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
OY 2641 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
DB 2641 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
OY 2701 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
DB 2701 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
```

D6	2761	ATGTGACACAGACGCGTGGTGTGGGATTCGTCACATTCACACACCTGTGTCAGTGGTTCTG	2760
QY	2761	CTACTCTACTGTTATTTTGTGAGAGGAACAAAAAGAAATCACTGTCACTTACTATGCGCC	2820
D6	2761	CTACTCTACTGTTATTTTGTGAGAGGAACAAAAAGAAATCACTGTCACTTACTATGCGCC	2820
QY	2821	CATTCACATTAAGAACCAATGGGGCTTTGGTGCCTGTTCAATGATTACATAAGATGTTGGAG	2880
D6	2821	CATTCACATTAAGAACCAATGGGGCTTTGGTGCCTGTTCAATGATTACATAAGATGTTGGAG	2880
QY	2881	CAGACACAGCAATAGAACCAACACCATCTACAGTCTCTGGCTGCTGATGATACACCCCTT	2940
D6	2881	CAGACACAGCAATAGAACCAACACCATCTACAGTCTCTGGCTGCTGATGATACACCCCTT	2940
QY	2941	TGCTGCTTTAATGTTTGTCATGTATGAAAGAAATACACTGCCATAATCTAATGTATAAAGT	3000
D6	2941	TGCTGCTTTAATGTTTGTCATGTATGAAAGAAATACACTGCCATAATCTAATGTATAAAGT	3000
QY	3001	CACGTGGGTCAGATCTAGAGCTTAAGTAAGACAGTCTGGGGTTTCAAAATGTTATATGTT	3060
D6	3001	CACGTGGGTCAGATCTAGAGCTTAAGTAAGACAGTCTGGGGTTTCAAAATGTTATATGTT	3060
QY	3061	CCATTAATAATGCAAAATAAACACCTCCATTAATTAATAAAAAAAAAAAAAAAAAAAAAA	3111
D6	3061	CCATTAATAATGCAAAATAAACACCTCCATTAATTAATAAAAAAAAAAAAAAAAAAAAAA	3111

RESULT 4	AX400072	AX400072	2286 bp	DNA	linear	FAT 06 JUN 2002
LOCUS	AX400072					
DEFINITION	Sequence	243	from Patent	WO0218424.		
ACCESSION	AX400072					
VERSION	AX400072.1		GI:21336495			

ORGANISM	REFERENCE
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1	
Tang, Y.T., Asundi, V., Xu, P., Xu, A. J., Fan, F., Zhang, D.,	
Wang, J.R., Zhao, Q. A., Wang, D., Liu, C., Drmanac, P. T., and Wehrman, T.	
Nucleic acids and polypeptides	
Patent: WO 0218424-A 243 07-MAR-2002;	
JOURNAL	

FEATURES	
source	Location/Qualifiers
	1. .2286
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
CDS	72. .1715

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CA033516.1"
/db_xref="GI:21336496"
/translation="MDIFSRNRKPLEVFAKVIQKIPAMIGQGLINFGIMHTVGF
SRINMFDGALHSTFTIRGVEYVPSKSLSDQITNTNINLEARIVSEFGMAVPDC
KNISKAESYLSHTIPDFTNCLINIMKGEQFIATSFNTYIRINPTLTLEKRVY
KRYAAVNLATSHPRHDEAGNVILNMGSTIVLEKQKRYVLEKIPATPEEKKQSGPMH
TEVCSIPSRSLSPSYSGFVGTENVVLELPDPELLIKMATYIPPMESCAIAF
HKEETVYTHIDORTPOVQIKFYTDAMVPHHNAVEDGQIVDAVDYDNLXYDL
FYELNINDEGERSLTSPTLPRFAPVPHNRNVEGNINLIKVASTATATIKFDD
VDCQPELTVEGLTPRYNVAHNGKQRYFPAAGVQMSPIPKIKIYDITLSLRKEE
DDCPRAEFLYPVPAKGADDDGVILSAISYIDQKPLFLIDAKSFTELAKASVDV
MHMDHGLFTDMQMDIKKQAASEEDRNASDCHGAPLT"

```

```
Query Match:      25.8%   Score: 904;    Pos: 67     Length: 2286,  
Best Local Similarity: 70.0%;   Prev. No.: 4 1-164;  
Matches: 1112; Conservative: 0; Mismatches: 470; Indels: 6; Gaps: 2.
```

OY	188	AGGTGTAAGGTGAGTTGGCTCTTCTGTTGTAAGAGAGTATCTGTTAAATGAGTCTGTAAGCA	247
DB	127	AAATGACAGGCAAGATTCCTAGTATGCTGTGTAAGAAATCTGTGTAATGAGTCTGGGCA	186
OY	248	TGCATCAATTAAGAGATACATAAATACCAACATGAGTTGTAAGGCTTGGTCTGTGTAAGCA	307
DB	187	TGCAGTAACTTGTGAGAGTCCAGATACCAACCACTGGTTGAGAGGAGCTTGGCTGTGCACA	246
OY	308	GCTTCACCTTTAAAAATGTTGAATTTTACTATACAGAAATAGTACCTCCCAAGTGCACAT	367
DB	247	GCTTCACCACTGAGAGAGCTGAATCTATATTAAGAGATTTGAAGAAATGAATAT	306
OY	368	ATATATTAATATAGAAATGAATATCTGTGTAAGTGTGTAAGTGAATGATGTTAT	427
DB	307	ACACACCAATATTAAGAGAGAGAGATTTGAGTGTCTGAGTTGAGCAATAGGCTATTC	366
OY	428	CGATCCATGCAAAATATATTTGCAAGATATTTATATATATATATATATATATATAT	487
DB	367	CGAGCACTTGTGAAGATTTATATATATATATATATATATATATATATATATATAT	426
OY	488	AGTTCAGGACAACTGCTGATCAACATTAATGAAGAAATGGGAGATATTTATGCTACCA	547
DB	427	ATTTCACAGACCACTGCTTGAATATGAATATATGAATGTGAGAGAGATTTGTAGAGACT	486
OY	548	GTSAGACTAACCTCAACAGAAATTTGATCCAGATCTGAGAGATCTAGATGAATAG	607
DB	487	CAGAGATTAATTAATTCAGAGAAATCAACCACTGAGATCTGAGAACTGAGAGGTTG	546
OY	608	ACTATAGTAAAT	667
DB	547	ATTATGCTAAAT	606
OY	668	GAATATATATCAAT	727
DB	607	GAATATATATCAAT	666
OY	728	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAAAGAGAAATCTGTTTAAACACC	784
DB	667	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	726
OY	785	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	844
DB	727	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	786
OY	845	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	904
DB	787	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	846
OY	905	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	964
DB	847	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	906
OY	965	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1024
DB	907	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	966
OY	1025	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1084
DB	967	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1026
OY	1085	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1144
DB	1027	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1086
OY	1145	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1204
DB	1087	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1146
OY	1205	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1264
DB	1147	TTAAGATCCCTCTCTGTACCAAG---AAAGAGAGAAAGAGAAATCTGTTTAAACACC	1206

DREKTYIHLIDORTKRPVPTKFTDPMVWVHHNNAEEDSCVFDIAYEDSLYL
 FYLANLNKDFEKPSPVPTVPPFAVLEHDKDAEVSNLVKVSTTATLAKEDH
 VYCOPEVLESELEPRTVYAVNPKPYRIFAEGVMSVPLILKYLTLKRSLSKSE
 ESCWPAEPLFVETPGAKDEJVLJLSAIVSTDPKLPFLILLDKSSTELARSVDAD
 MHLDHGLFIPDADNNNAVKUTPAETJUEVENSHPIDLPAPLHSENDPFAHSGSSL
 "

BASE COUNT 627 a 512 c 554 g 424 t

ORIGIN

Query Match 25.0%; Score 777.4; DB 10; Length 2117;
 Best Local Similarity 68.8%; Pred. No. 2,5e-158;
 Matches 1098; Conservative 0; Mismatches 491; Indels 6; Gaps 2;

121 CACAGAGAGGAGTGGAGACATATTTAACAGAAAGAGAGACATCCAGAGCCATA 180
 167 CTCACCGGTTCCATGGAGTAAATATTGGCCAGAAATAGAAAGACACCTGGACCGATT 226
 181 AAGAGTGAAGTGCMAAGTCACTTGGCCACTTGGTTCAGAGGCTACTCTCGGAATGAC 240
 227 CAGGGCCAAAGTGCACAGGACGATTTCCAGCATGCGTGCAGGAGCCCTGCTCGAAGCGG 286
 241 CCAGGATGCACACAAATAGGGGACCTAATACACACCTGGTTGATGGCTGGCTG 300
 287 CCGGAGATGCACACAGTGGAGAGAGCAATGACCAACCATGGTTGATGGCTGGCTG 346
 301 CTCACAGCTCAAGCTTTAAATAAGTAACTTACAGAAAGTAACTAGTACCTCGAAGT 360
 347 CTCACAGCTTCTCCATCGAGATGGAGAGGCTCTCTACAGAGCAATACCTCGAGAGT 406
 361 GACACATACACATGCAATATAGAGCAAAACCAATGAGAGTCTGAGTGGAGCAATG 420
 407 GACACATACATGCAATATAGAGCAAAACCAATGAGAGTCTGAGTGGAGCAATG 466
 421 GCTATCGGAGTCCATGCAAAACATATTGGCAGGCTTCTCTACTTATCTACACCC 480
 467 GCTATCGGAGTCCATGCAAAACATATTGGCAGGCTTCTCTACTTATCTACACCC 526
 481 ATTCTGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATAT 540
 527 ATCCCGAGCTTCACAGCACTGCTGATGACACATATGAAACAGGAGATGATATAT 586
 541 GCTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATAT 600
 587 GCAACACGAGTTCACAGCACTGCTGATGACACATATGAAACAGGAGATGATATAT 646
 601 AAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 660
 647 AAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 706
 661 AAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 720
 707 GAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 766
 721 GCTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATAT 777
 767 GCTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATAT 826
 778 AAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 837
 827 AAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 886
 838 CACAGCTTGAAGTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 897
 887 CACAGCTTGAAGTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 946
 898 ATTGTAAAGTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 957
 947 ATTGTAAAGTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1006
 958 CATAAGAGATAGAGCTGCTGATGACACATATGAAACAGGAGATGATATATATAT 1017
 1007 GACAGAGATAGAGCTGCTGATGACACATATGAAACAGGAGATGATATATATAT 1066

1018 ACAGATTTTATATGATGCTTTTATATATATATATATATATATATATATATATAT 1077
 1067 ACCAAGTCTTAT 1126
 1078 GACAGCTTGAAGTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1137
 1127 GCTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1186
 1138 TTAAGAAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCA 1197
 1187 CTGAGCACTTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCA 1246
 1198 AAGGTAGCTACAGCACTGCTGATGACACATATGAAACAGGAGATGATATATATATAT 1257
 1247 AGAGGCTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGA 1306
 1258 AAGGTAGCTACAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCA 1314
 1307 AAGGTAGCTACAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCA 1366
 1315 CCGTAAAT 1374
 1367 CCGTAAAT 1426
 1375 AATATAGCTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGA 1434
 1427 CCGTAAAT 1486
 1435 AATATAGCTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGA 1494
 1487 AATATAGCTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGA 1546
 1495 GAGGTAGCTACAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGTGA 1554
 1547 GAGGTAGCTACAGCACTTGAAGTGAAGCACTTGAAGTGAAGCACTTGAAGTGAAGTGA 1606
 1555 TGTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1614
 1607 GCTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1666
 1615 TGTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1674
 1667 TGTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1726
 1675 TGTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1709
 1727 TGTATCGAGTTCACGAGCACTGCTGATGACACATATGAAACAGGAGATGATATATAT 1761

RESULT 8
 AF294899
 LOCUS
 DEFINITION
 complete cds.
 AF294899
 AF294899
 1
 10242317
 ACCESSION
 VEPSTIN
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Mus musculus.
 Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
 Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Mus.
 1 (bases 1 to 2119)
 YAN, W., JANG, G.-F., HASELEER, F., ESUMI, N., CHANG, J., KERRIGAN, M.,
 CAMPOCHIARO, M., CAMPOCHIARO, P., PALCZEWSKI, K. and ZACK, D.J.,
 Cloning and characterization of a human beta₂ carotene
 15,15'-dioxygenase that is highly expressed in the retinal pigment
 epithelium
 Unpublished
 JOURNAL
 2 (bases 1 to 2119)
 YAN, W., JANG, G.-F., HASELEER, F., ESUMI, N., CHANG, J., KERRIGAN, M.,
 CAMPOCHIARO, M., CAMPOCHIARO, P., PALCZEWSKI, K. and ZACK, D.J.,
 Direct Submission
 Submitted (09-AUG-2000) Ophthalmology, Johns Hopkins University,

600 N. Wolfe St., Baltimore, MD 21287, USA
 Location/Qualifiers
 source

1..2119
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone_image="2192191"
 /issue_type="kidney"
 /note="Sugano mouse kidney mklia"
 1..2119
 /gene="bedo"
 /gene="bedo"
 180..1880
 /EC_number="1.13.11.21"
 /codon_start=1
 /product="beta, beta-carotene 15,15'-dioxygenase"
 /protein_id="AA015381.1"
 /db_xref="GI:10242318"
 /translation="MEIIFGQNKKEQLEPYQAVTGSIPAMLOGTLLRNQPKMHTVGE
 SKYHMFEDGLALHSFIRDEGVYRSKYLOSPTYIANIEANRLVYSEGTMAVDPDC
 KNIFSKAFSYLSHRIIDPFDNCLINIKGCEDEYATTEHYIKRIDPTELEKYD
 RKYAVVNLATSHPHXPDAGNVLNMGTSVYKGRKYIETIPATVPDSKKKSPYKH
 AEVFCISSRSLSPTSHSGVTENVVLEDPFKDLIKMATVATMRGVSMSKSF
 DREKTYIHIDORTKRPVPKLEYTDPMVVFHVNAYEEDGCVLPVIAVEDSIIYL
 FYLANLNKDEPEKSLTSPILRFAVLDVDAEYGSMLVYSSITATLEKDG
 VYCOPEVLYEGLEPLRTNVAENGPRYIEAAEYQMSPPVTKILKDYDILKSLKSE
 ESCWPAELFVPTGAKDEDDGVLSAIVSDPQKLPFLILIDAKSTELARASVAD
 MHLHGLFIPDADWMAVKOIPATQEVENSHPIDPTABELSHENDPTAGHGSSL"

BASE COUNT 630 a 513 c 555 g 421 t
 ORIGIN

Query Match 24.9%; Score 775.8; DB 10; Length 2119,
 Best Local Similarity 68.8%; Pred. No. 5.5e-158;
 Matches 1097; Conservative 0; Mismatches 492; Indels 6; Gaps 2;

121 CACAGAGAGCATTGAGACATATTTAACGAAACAAAGAGCCTCCAGGCCATA 180
 168 CTCACCGGTTCCATGAGATATATTTGGCCGAGATTAAGAAAGACAGCTGGAGCGAGTT 227
 181 AAATTCAGGTGCAAGGTGAGTTCCTGACTTGTGTGCAAGGGGACTTCTCGAATAATGCC 240
 228 CAGCTCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 287
 241 CCAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 288 CCGGAGATGACACAGATGCGAGAGAGACACCATTTGTTGATGGCGCCCTT 347
 301 CTGACAGCTTCAGCTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATG 360
 348 CTCACAGCTTTCTCATCAGATGAGATGAGATGATGATGATGATGATGATGATGATGATGATG 407
 361 GATATATCAACTGATATATATAGAGCAAAATGATGATGATGATGATGATGATGATGATGATGATG 420
 408 GACAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 467
 421 GCTATCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
 468 GCGTACCGGACCCCTGCAAAACATCTTTCCAAAGCTTTCTCTACTTGTCTGACACC 527
 481 ATTCTGATTCAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 528 ATCTCGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 587
 541 GCTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 588 GCAATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 647
 601 AAGCTAGACTGACAGCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 648 AAGCTATATTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 707
 661 AGTCTGGAATATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720

708 CAGGCTGGAATATCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 747
 721 GTTCTCTTTAAGATCCTTCTCTCTGATGACAGCAAAAGCAAGAA--GAATTTCTTTT 777
 768 GATATATTAAGATCCTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 827
 778 AATATATTAAGATCCTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 847
 828 AAGCAGCGAAGTTTCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 887
 838 CACAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 897
 888 CAAATATTAAGATCCTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 947
 948 ATCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1007
 958 CATAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1017
 1008 GACAGGAGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1067
 1018 ACCAGATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1077
 1068 ACCAGATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1127
 1078 GCGCAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1137
 1128 GCGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1187
 1138 TTAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1197
 1188 CTGATG 1247
 1198 AAGCCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1257
 1248 AGGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1307
 1258 AAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1314
 1308 AAGGATG 1367
 1368 CCGGAGATG 1427
 1375 AAATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1434
 1428 CCAATG 1487
 1435 AAATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1494
 1488 AATATG 1547
 1495 GAGCCTATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1554
 1548 GAGCCTATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1607
 1555 TGTGTGATG 1614
 1608 GCTATG 1667
 1615 TTAATAAATG 1674
 1668 TTAATAAATG 1727
 1675 TTAATAAATG 1709
 1728 TTAATAAATG 1762

RESULT 9

AF271298 2137 bp mRNA linear ROD 27 FEB 2001
 LOCUS AF271298 Mus musculus beta-carotene 15,15'-dioxygenase mRNA, complete cds.
 DEFINITION AF271298
 ACCESSION AF271298
 VERSION AF271298.1 GI:11420969
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Gradinata; Vertebrata; Eutelestomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2137)
 Redmond, T. M., Gentileman, S., Duncan, T., Yu, S., Wiggert, B., Gantt, E. and Cunningham, F. X., Jr. Identification, expression, and substrate specificity of a mammalian beta-carotene 15,15'-dioxygenase
 J Biol Chem 276 (9), 6560-6565 (2001)
 JOURNAL 2 (bases 1 to 2137)
 MEDLINE 2125869
 PUBMED 11092891
 REFERENCE 2 (bases 1 to 2137)
 Redmond, T. M. and Gentileman, S. Direct Submission
 Submitted (22-MAY-2000) LMGMR, NPI, NIH, RIG 6, RM 339, 6 Center Dr., MSC 2740, Bethesda, MD 20892-2740, USA
 JOURNAL Location/Qualifiers
 TITLE 1. 2137
 FEATURES
 source
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 198..1848
 /codon_start=1
 /product="beta-carotene 15,15'-dioxygenase"
 /protein_id="FAG33982.1"
 /db_xref="GI:11320970"
 /translation="MELTFGQKKKQLEPVQAKVTGSIPLALQCTLLRNGPQHTVGE
 SKYNNHFDGLALHSESLPGEVEFVPSVYLSSTYIANIFANILVSEFQIMAYPLP
 KNIFSKASLYLSHTIPDITDNLINIMKGEDFATITINIRKIDQLETKVDY
 RYVAVNATLSHPHYDFAGNVLNM;21SVLEK;PIKYYVFKIPALVPLSKKSKSPVH
 AVEPCSSISRLSPSYHSGVTEVNYEVEPEKIDILKMATAYKRGVSNMSCE
 DREDKTYHIIDQFRKRPVPTKFTYIPVAVHNAVHEGSLVPLVIAHNSGNSCE
 FYLANLNKDEPKSKLISVPTLRRKRAVLHYDKDAEVSNLVYKVSSTTATLAKKDDH
 VYQPPVLEFEGLEIPRIYATVNGKRPYHFAEYVQMSVPLKILKYHITLSSISKWSF
 ESCWPAEPLFVPTGAKDEEDVYIIISAVISTDPQIPILILDKSSTELARAVDAD
 MHLDHGFLTPDAIMWNAVKLTPAEILQEVENSHPILPLIAELSHSEMTFA;385SSSE
 *
 BASE COUNT 633 a 518 c 556 g 424 t 1 other s
 ORIGIN
 Query Match 24.98: Score 775.8; DB 10; Length 2137;
 Best Local Similarity 68.88; P-val 5.4e-118;
 Matches 1097; Conservative 0; Mismatches 492; Indels 6; Gaps 2;

Db 486 GACTACCGGATCCTTGA AAAAACATGCTTTCTTACTTGCTTCACACC 545
 Qy 481 ATTTCGAGTGTGAGGAAACAGGCGAGAACATATGAAGAAATGAGGATATATAT 540
 Db 546 ATCCCGSATTACAGAAATATCTGTATATGAATATATGAATATGAATATGAAT 605
 Qy 541 GTTACGATGAGTAACTTGTATGAGAAATATATGATATATGATATGATATGAT 600
 Db 606 GCAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
 Qy 601 AAGGTAGACTACAGGAAATAGTACTGTAATCTGGCAATCTGACACCACTATGAC 660
 Db 666 AAGGTATATATCCGAAATATATATATATATATATATATATATATATATATAT 725
 Qy 661 AGTCTGGAAT 720
 Db 726 GAGGTAT 785
 Qy 721 GTTCTCTTAAAGATGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 777
 Db 786 GTTAT 845
 Qy 778 AAACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 847
 Db 846 AAGCATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 905
 Qy 838 CACAGCTTTGGAATACAGCAAAATATATATATATATATATATATATATATATAT 897
 Db 906 CAAAT 965
 Qy 898 ATTGTGAAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 957
 Db 966 ATCTCTAATATGCGCCACCCACATACATGAGGAGGAGGAGGAGGAGGAGGAGG 1025
 Qy 958 CATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1017
 Db 1026 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1085
 Qy 1018 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1077
 Db 1086 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1145
 Qy 1078 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1137
 Db 1146 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1205
 Qy 1138 TTAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1197
 Db 1206 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1265
 Qy 1198 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1257
 Db 1266 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1325
 Qy 1258 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1314
 Db 1326 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1385
 Qy 1315 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1374
 Db 1386 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1445
 Qy 1375 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1434
 Db 1446 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1505
 Qy 1435 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1494
 Db 1506 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1565
 Qy 1495 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1554
 Db 1566 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1625

[illegible][illegible]

Db	1166	GAATTCATGCGCTTTGAGAACTAGGGCTGTGTTTAAATTCATTTGGTGGCTGCGCAAAATAT	1226
QY	1114	AATAGCGTTAGCATATGTTTACTTAAAAAAACCT-----GGACAAAGACTTTGAA	1166
Db	1226	GTAAATATCTTGAATTTTACTTTATATATATCTTAAAGAAAGCTTAAAGAAAGAGCTTAAAT	1285
QY	1165	GTCATACACAGAGCTTACCTGCAATGCCAAGCTGCGAAAGAGCTTTGTGGGCTGCGAGTAT	1226
Db	1286	CAGTCTCATTAATTCAGACAGCCAAATCTTGGCTGGAGAGTTGTTTGGCTTTAAATATGC	1344
QY	1225	GATCAAGATGCGAAGATAGTTTCTAATTTAACTCAACCTCC---AACCTCCGACTGCT	1285
Db	1346	AGTTTAAATGAGCGCTGAGAGAGAGCAAGCTGAGGCTATTTGCTATCTTAACTTAACTGCT	1406
QY	1282	GTAAGAGAAAAAGATGCGAGCATCTATGTCACACTGCAATATTATATCT-----	1326
Db	1406	GTCATAAATAGGCTGATATGAAATATCTGAGTCTCTATATAAATCTCAACTAGAGAGAGCTTA	1465
QY	1330	-----GAAGCATAGAACTGCTGCTGTCTACATATATACATCAATAGGTAAGAAATATCT	1480
Db	1466	GAAAAATAAAGAGAGGCTATTTGAAATTTGCTATATATATATATATATATATATATATATAT	1525
QY	1381	AATATATGTTATGCAACAAAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1440
Db	1526	TATCATTTCTTTTATGCGCTGTGGCTTTGGCATTTAGTGGGAGATTTCTGATCAAGGTT	1585
QY	1441	AATGTCCAAATAAAT	1500
Db	1586	GATGTGTCAATTAACATCATTAAGGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1645
QY	1501	ATCTTTGTGTCGCAATGCGGATGTAAGAGAAAGATTAAGATGTGTGTGTGTGTGTGTGTGT	1566
Db	1646	GTTTGTGTTTTCACAT	1705
QY	1561	GTGATGTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1620
Db	1706	ATCATGCCACACCATATGAAAGCAATTTTCTGATTTTGGATGGCAAGAACTTGA	1765
QY	1621	GAATTTGGCGGCGCTTACATTAATGTAATTAATGATTTGATTTGTAAGAGAGAGAGAG	1680
Db	1766	GAGTGTGTTGAG	1825
QY	1681	CC 1682	
Db	1826	CC 1827	
RESULT 14			
AF276432	2597 bp	mRNA	linear
LOCUS			PR1.16-ARG-2001
DEFINITION			Homo sapiens putative carotene dioxygenase mRNA, complete cds.
ACCESSION	AF276432		
VERSION	AF276432.1	GI:14582264	
KEYWORDS			
SOURCE			
ORGANISM			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
FEATURES			
source			

[illegible]


```

Db      424 ATCTGAGATTGGGACACTGCTCTCTCCGATCCATGCAAGATGTTTGGACCTTTG 483
QY      454 AAGCATTCTCATCTACTATCTGACACCATCTCTGAGTTCACGGACCACTGCTGATCAAC 513
Db      484 ATGTCCAGGTTGGAGCTGCTGTAAGCTGACGCCATGACTGACCAATACTATATGTCAC 543
QY      514 ATTAGAAGAACTGGGGATGATTATATGCTACCGCTGAGACTAGACTAGCTTACATGACAAAAT 573
Db      544 TAGTGGCGGTACAAAGGGGATTAATCTGACCTCTGACCTGAGACCAACTTATGAATTAAGTG 603
QY      574 GATCCACAGCTCTGGAGACACATAGATTAAGTATGACTACAGCAATATATGAGCTGAAC 633
Db      604 GACATTTGAACCTCTGGAAAAAATGAAAAAGGTATATGAGCAATATATGCTGAGAA 663
QY      634 TTGCTAACTTTCTTACCTTATGATATGCTGCTGAAATATTTGCAATGAGTACTGTA 693
Db      664 GATCTAACTGTAATCTCTCATTAATGAACTGAACTGAACTAACTAACTAACTGAACTG 723
QY      694 ATGTGTTAAGGAGAAATCTGTTTAAACCTGAGAGTATGCTGATCTCTGACAGAA 753
Db      724 TTTC-----GGCCATATGTTTCTCTCATTAAGCTTATGCGGTTCTCTCCAGAG 771
QY      754 AAACAAAAGAGAAATCTGTTTAAACCTGAGAGTATGCTGATCTCTGAC 813
Db      772 AAGTGGACCTTGAGGAGAACTATGCTGATGATGCTGATGCTGATGCTGATGCTGATG 831
QY      814 TCCCTGCTCAGCAGCTACTACACAGCTCTGAGATGACAGAAAAATATATGCTG 873
Db      832 GAGAAAGGAGAACTCTTACTACTACATAGCTTTGAAAGGACAAAGAACTATATATTTTC 891
QY      874 ATAGAGCAGCATTTAAGCTGATATGCTGAACTGCACTGCTACATCCGAGGTG 933
Db      892 ATTGACAACTCTTAAAGATGAACTGTTGAAATGCTGAAATGCTGAAATGCTGAAATG 951
QY      934 AACTGGGCTTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 993
Db      952 GCTTTTCTGATGGATGAGCTGAGAACCCAGCTGATATGCGGCTTCTTCTTCTTCTTCT 1011
QY      994 AGAAAGAGAAAGAAAGATATCTACCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1053
Db      1012 AAAGCAGCTGAGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1071
QY      1054 CATATTAATGTTTCTGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1104
Db      1072 CAATCAATGCTTCTGAGAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1131
QY      1105 TACAGAGCACTAGCTTCTGATGATATGTTTACTTAAAGAACTGAGCAAGCTTGA 1164
Db      1132 GGAAGAACCTAGAAAGTTTACCACTTACAGAACTGAGAGAGAGAGAGAGAGAGAGAG 1191
QY      1165 GTGAACTAAGCTTACCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1224
Db      1192 GAGCTCTATATTTAGTAACTAAATTTTCTCTCTGAGAGAGAGAGAGAGAGAGAGAG 1251
QY      1225 GATTAAGCTATGAGAGATGCTTCTTATTTACTTAAATTTGCTGCTGCTGCTGCTGCTG 1281
Db      1252 AGTTGATGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
QY      1282 GTAAAGAAAAAGATGCTGATCTTATTTCTAACTGAGAAATATATGAGAGAGAGAGAG 1341
Db      1312 GTGAAATAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
QY      1342 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1401
Db      1372 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
QY      1402 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
Db      1432 TATCATTTCTTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491
QY      1459 GTACTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518

```

```

Db      1452 GATGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
QY      1519 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
Db      1552 GAAAGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
QY      1579 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
Db      1612 GAAAGCAATTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
QY      1639 GTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1682
Db      1672 GAAAGCAATTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1715

```

Search completed: July 15, 2003, 20:05:36
 Job time : 7877 secs

0 0 0 0 0 0 0 0 0 0 0 0

VERSION BC02259.1 GI:18490381
 KEYWORDS HTC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini, Hominoidea; Homo;
 1 (bases 1 to 974)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20832-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs@mail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: The I M A G E Consortium (ILNL)
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.simg.stanford.edu>
 Contact: (Dickson, Mark) mdickson@stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 P. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 37 Row: 3 Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA 91: 8393364
 This clone has the following problem: Incomplete processing.
 Location/Qualifiers

FEATURES

source

1..974

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4734856"
 /tissue_type="Liver"
 /clone_lib="NIH-MGC_76"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 BASE COUNT 310 a 230 c 249 g 185 t
 ORIGIN

Query Match 9.4%; Score 293.4; DB 11; Length 974;
 Best Local Similarity 75.8%; Pred No 47e-42;
 Matches 363; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

128 GAGCGATGAGACAAATATTACGAAAGCAAGAGCATCCAGAGCCCATAAAGCTG 187
 214 CAGCAATGATATATATTTGGCGAATAGGAAAGACGCTGGAGCCTGTGAGGCCA 273
 188 AGGTGCAAGTCACTTCCCACTGGTGGCAAGGGGACTCTCCGGAATGGCCAGGA 247
 274 AATTACAGCTCAAGATTCACATGCTGCAAGGACCTGCGCAATGCGCGTGGGA 333
 248 TGCACACATTCGGGACACATAATACACAGCAGCTTGATGCTGCTGCTGCTGCTG 307
 314 TGCACATAGTTGGGATTCAGATACCACTTGGTGGAGCGGCGCTGGCTCCCTCA 393
 308 GTTTCATTTAAATAGTGAAGTTTACTACAGAACTAGTCCGCAAGTGCACAC 367
 394 GCTTCACGATCAGACGCGTGAAGTCTATTACAGAGCAATACCTGAGAGGAGTAC 453
 368 ACAACTGCATATATGAGCAAAACCGAATGCTGTGTGATGATTTGAGACCATGGCTTATC 427
 454 ACAACACCAATATATGAGCAAAACCGAATGCTGTGTGATGATTTGAGACCATGGCTTATC 513
 428 CGGATCCATGCAAAATATATTTCCGCAAGCATTTGATCTATCTACACATTTCTG 487
 514 CGGACCCCTGCAAAATATTTTCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 573
 488 AGTTCAGGATCACTGATGATTAATATATGAAAGATGAGATATATATATATATGATA 547

Db 574 ATTTACCGGACAGACTGCTGATCAACATATATGAGAGAGATTTTACGCTGACT 643
 548 GTGAGACTAATCTATCAGAAAATGATCCAGACATCTGAGACATGATATAGCTA 606
 634 CAGAGACCAATATATATGAGAAAATCAACCCAGAGACATCTGGAATCTGAGAGAGCTA 692

RESULT 5

AM044715

LOCUS um13104.y1 Sugano mouse kidney mKia Mus musculus cDNA clone
 DEFINITION IMAGE:2192191.5' similar to TR:Q16518 Q16518 RETINAL PIGMENT
 EPITHELIAL-SPECIFIC 61 KDA PROTEIN ; mRNA sequence.

ACCESSION

AM044715

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Marla M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through ILNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:1004643

Seq primer: custom primer used

High quality sequence stop: 473.

Location/Qualifiers

1..642

/organism="Mus musculus"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:2192191"

/clone_lib="Sugano mouse kidney mKia"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pMR18-P13; Site: 1; Inset 1
 (CAGCTGCTG); site 2: primer 1 (CAGCAATG); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGCGCTTTTCTTTTCTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor (TGTGGCTCTACTGG), digested
 and cloned into distinct DraIII sites of the pMR18 P13
 vector (5' site CAGCTGCTG, 3' site CAGCAATG). Most should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments < 1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTTCTGCTCTTAAAGATTTG3' and 3' end
 primer CGACCTGCACTGAGACCA."

BASE COUNT 186 a 169 c 168 g 119 t

ORIGIN

Query Match 9.2%; Score 285; DB 10; Length 642;
 Best Local Similarity 74.2%; Pred. No. 1.6e-40;
 Matches 360; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

121 CAGCAATGATATATATTTGGCGAATAGGAAAGACGCTGGAGCCTGTGAGGCCA 180
 157 CAGCAATGATATATATTTGGCGAATAGGAAAGACGCTGGAGCCTGTGAGGCCA 216


```
Db      643 AAAAATTTACAGTGGCTACACATTATTAAAGATGCT-----CCACTGAAACCA 693
QY      760 AAGAGAAATCTTG1111AAACACUUGAAGTAGATGTCACATCCCTTCGGTCCG 819
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      694 GACAGAGAAAGATCCCAATAAACCAAGTCASAG11GTTGSCAG11CCCTGAGTAGTGT 753
QY      820 CTCCAACCAAGCTACTACGACAGCTTTTGTATATCAGAGAAATATATGTTGTCATAGAG 879
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      754 TTTAAGTCATTTATATATACACATTTTATTTGACTCCTAACCTATATCTTTTGTGAG 813
QY      880 GAGCCATTTAACTGATATATGCA 905
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      814 ACTCCAGTCAAAATTAACCTTTTCA 839
```

Search completed: July 15, 2003, 21:28:40
Job time : 4173 secs



513 CATTATGAAAGTGGGATGATATATATGCTACCGAGAGTAACTGATGAGAAAT 572
 442 GATACCCAGTAGAGGAAATATATATGCTGAGAGAGAGAGAGTAAAGAT 491
 573 TGATGACAGAGTGTGAGAGAGTATATAGTATAGTATAGAGAGAGATATAGTATAA 632
 492 TAATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTAT 551
 633 CTGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 692
 552 TGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 611
 693 AATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 752
 612 CTGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 662
 753 AAGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 812
 663 CTGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 722
 813 CTGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 872
 723 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 782
 873 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 929
 783 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 842
 940 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 999
 843 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 902
 990 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1049
 903 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 962
 1050 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1106
 963 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1022
 1107 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1166
 1023 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1082
 1167 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1220
 1083 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1142
 1221 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1277
 1143 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1202
 1278 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1327
 1203 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1262
 1348 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1385
 1243 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1322
 1386 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1445
 1323 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1382
 1446 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1505
 1383 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1442
 1506 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1565
 1443 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1502

QY 1566 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1622
 Db 1503 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1562
 QY 1643 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1640
 Db 1563 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1620
 RESULT 2
 US-09-645-370-1
 ? Sequence 1, Application US/09/645,370
 ? Patent No. 6428958
 ? GENERAL INFORMATION:
 ? APPLICANT: Aguilera, Gustavo D.
 ? APPLICANT: Aguilera, Gustavo M.
 ? APPLICANT: Ray, Kunal
 ? TITLE OF INVENTION: IDENTIFICATION OF CONSENTAL STATIONARY NIGHT HUMIDUS
 ? FILE REFERENCE: 19601/2481
 ? CURRENT APPLICATION NUMBER: US/09/645, 370
 ? PRIOR FILING DATE: 2000-08-24
 ? PRIOR APPLICATION NUMBER: 09/385,259
 ? PRIOR FILING DATE: 1999-08-30
 ? PRIOR APPLICATION NUMBER: 60/103,219
 ? PRIOR FILING DATE: 1998-10-06
 ? NUMBER OF SEQ ID NOS: 8
 ? SEQUENCE: Patent No. 6428958
 ? SEQ ID NO: 1
 ? LENGTH: 1724
 ? TYPE: DNA
 ? ORGANISM: Canis familiaris
 US-09-645-370-1
 Query Match 8.5%; Score 264.8; DB 4; Length 1724;
 Host Local Similarity 52.1%; Pct: No. 5,5e-61;
 Matches 811; Conservative 0; Mismatches 702; Indels 45; Gaps 8;
 QY 153 AAGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 212
 Db 78 AAGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 147
 QY 213 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 272
 Db 138 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 197
 QY 273 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 342
 Db 198 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 257
 QY 333 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 392
 Db 258 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 317
 QY 393 AAGAGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 452
 Db 318 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 477
 QY 453 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 512
 Db 378 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 441
 QY 513 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 572
 Db 432 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 491
 QY 573 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 632
 Db 492 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 551
 QY 633 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 692
 Db 552 GATGAGAGAGTGTGAGAGAGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 611

QY	1847	TCCTTTAA	1855
Db	1973	GTCATGTTA	1965

RESULT 13

```

US-09-124-523-28/c
Sequence 28, Application US/09124523
Patent No. 6395,660

GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HUBWITZ & THIBEAULT
STREET: High Street Tower . 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #10, Version #1 to
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 2307 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-124-523 28

```

Query Match	1.38	Score 43.8	CR 4	Length 2307
-------------	------	------------	------	-------------

QY	1787	AAGCAGACCGCATATTAAGTCAACACAAAGAAGACATAGCATATTCCTTTAAATAAAGATATAAA	1846
Dd	2633	ATCTAACATCATGATATTATCAAATATATGTGTAACTGGTGCTTATGATGATGATATAT	1974
QY	1847	TCCCTTTAA	1855
Dd	1973	GTCATGTTA	1965

RESULT 14

US-08-980-060-3

; Sequence 3, Application US/2004/0100000
; Patent No. 5965421

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match: 1.38; Score 41.2; MB 2; Length 3459;

[illegible]

RESULT 15

US-09-007-185-3
Sequence 3, Application US/09307185
Patent No. 6,22,019
GENERAL INFORMATION:-
APPLICANT: NI JIAN
APPLICANT: FENG, PING
APPLICANT: MOJLO, MARIA
APPLICANT: DIXIT, VISHVA M.
TITLE OF INVENTION: HUMAN IRAK-2
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/307,185
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/980,060
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3459 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOL.PCULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..1908
US-09-307-185-3

Query Match 1.3%, Score 41.2, DB 4, Length 3459;
Best Local Similarity 53.0%; Pred. No. 0.39;
Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

OY	2946	TCTTATGCTTGCATGATGAGATACACTGCCTAATCTATGTTAAAGTCAC	3005
DB	3282	TATTCAGTCGCTATATATCTGATAAAGCATGTTGCCAATCTTCTGTGAACAGATGCTA	3341
OY	3006	GCGTCAGATCTAGACCTTAAGTAAGCAGCTGGGGTTTCAATGTTATATGTTCCATA	3065
DB	3342	TTTAAATTCACCTGAGCAAAATATCTATTAAAGTATCTATAGTAATTTCTTTTATA	3401
OY	3066	AAATGGAATTAACACCTCCATATATAAAAAAAAAAAAAAAAAAAAAA	3111
DB	3402	TAAATAAATATATTTTGTAAAGTGAAGAAAAAAAAAAAAAAAAAAAA	3447

Search completed: July 15, 2003, 20:18:55
Job time : 140 secs

19	41.2	1.3	3459	10	MS-09-773-753-3	56
----	------	-----	------	----	-----------------	----

Description

Result No.	Score	Query Match	length	DB	ID	Description
1	3111	100.0	3111	9	US-10-053-192.2	Sequence 2, Appli
2	46.4	1.5	3037	4	US-10-172-086.84	Sequence 89, Appl
3	45	1.4	5689	9	US-10-239-676-89	Sequence 89, Appl
4	44.2	1.4	2000	9	US-09-938-842A-4412	Sequence 4912, Appl
5	44.2	1.4	2000	10	US-09-947-376-39	Sequence 4912, Appl
6	43.6	1.4	357	4	US-10-060-046.2845	Sequence 39, Appl
7	43.2	1.4	588	9	US-09-871-161-408	Sequence 2800, Ap
8	43.2	1.4	11250	9	US-10-239-676-20	Sequence 406, Appl
9	42.8	1.4	478	4	US-05-918-995-2852	Sequence 20, Appl
10	42.2	1.4	6046	9	US-10-239-676-16	Sequence 23852. A
11	42.2	1.4	8966	9	US-10-239-676-16	Sequence 16, Appl
12	42	1.4	1410	9	US-10-099-278.81	Sequence 139, Appl
13	41.8	1.3	2804	4	US-09-938-842A-1175	Sequence 81, Appl
14	41.8	1.3	5285	10	US-09-925-301-102	Sequence 4170, Ap
15	41.8	1.3	5650	9	US-10-038-001-1	Sequence 102, Appl
16	41.8	1.3	70768	9	US-10-038-001-2	Sequence 1, Appl1
17	41.6	1.3	8693	9	US-10-135-323-13	Sequence 2, Appl1
18	41.6	1.3	8693	9	US-10-172-086.38	Sequence 13, Appl
19	41.2	1.3	3459	10	US-09-774-755-3	Sequence 38, Appl
						Sequence 3, Appl

C	20	41.2	1.3	4598	9	HS-10-144-020-27	Sequence 28, Appl
C	21	41.2	1.3	15732	9	HS-10-230-676-06	Sequence 95, Appl
C	22	41.2	1.3	133121	9	US-10-125-523-71	Sequence 71, Appl
C	23	41.2	1.3	198285	10	US-09-180-197-3814	Sequence 3814, A
C	24	41	1.3	696	10	US-09-770-693-9	Sequence 9, Appl
C	25	41	1.3	17848	9	US-10-239-676-28	Sequence 28, Appl
C	26	40.8	1.3	11812	9	HS-10-240-676-204	Sequence 209, Appl
C	27	40.8	1.3	84373	10	US-09-263-665-1	Sequence 1, Appl
C	28	40.6	1.3	858	9	US-10-198-846-7505	Sequence 7505, A
C	29	40.6	1.3	2000	9	US-09-148-846-2791	Sequence 2791, A
C	30	40.6	1.3	2494	9	US-10-152-661-556	Sequence 556, Appl
C	31	40.6	1.3	2494	9	US-09-856-050A-556	Sequence 556, Appl
C	32	40.6	1.3	3436	9	US-09-740-676-22-3	Sequence 3, Appl
C	33	40.6	1.3	3436	9	HS-10-141-953-3	Sequence 3, Appl
C	34	40.6	1.3	3436	10	US-09-789-561-12	Sequence 12, Appl
C	35	40.6	1.3	127197	9	US-09-754-853A-1	Sequence 1, Appl
C	36	40.4	1.3	521	9	US-09-444-707-234	Sequence 234, Appl
C	37	40.2	1.3	1492	9	US-10-180-375-29	Sequence 29, Appl
C	38	40.2	1.3	1741	10	HS-09-855-548A-5	Sequence 5, Appl
C	39	40.2	1.3	5502	9	US-09-808-602-3	Sequence 3, Appl
C	40	40.2	1.3	5502	9	US-09-965-546-17	Sequence 17, Appl
C	41	40.2	1.3	5502	9	US-09-965-545-17	Sequence 17, Appl
C	42	40.2	1.3	5502	9	US-09-965-512-17	Sequence 17, Appl
C	43	40.2	1.3	5502	9	US-09-900-194-3	Sequence 3, Appl
C	44	40.2	1.3	513505	9	US-09-754-853A-4	Sequence 4, Appl
C	45	40	1.3	313	9	HS-09-148-846-2791	Sequence 2791, Appl

ALIGNMENTS

```

RESULT 1
US 10 053 192-2
Sequence 2: Application US/10053192
Publication NO: US/2004/0087336A1
GENERAL INFORMATION:
APPLICANT: BACHMANN, Heinrich
APPLICANT: BRUNGER, Roland
APPLICANT: FRIEDLEIN, Arno M
APPLICANT: WIRTH, Gabriele M
APPLICANT: WOODSON, Wolf-Dietrich
APPLICANT: WYSS, Adrian
APPLICANT: WYSS, Markus
TITLE OF INVENTION: BETA-DEFENSIN 15, 15'-POLYMERANES, NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
FILE REFERENCE: B, H-CAPOTENF 15, 15'-POLYMERANES, ...
CURRENT APPLICANT ID NUMBER: US/10053192
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 103382.0
PRIOR FILING DATE: 1998-02-22
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 2
? LENGTH: 3111
? TYPE: DNA
? ORGANISM: CHICKEN
US-10-053-192-2

Query Match      100.0%   Score 3111;   NR 9;   Length 3111;
Host Local Similarity 100.0%;   Pred No 0;
Matches 3111;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0

1  CGAGACACATAGACAGCGCGGACAGCGGAGATCATCTCTGTATACAGCAAG 60
|||||
1  CGAGACACATAGACAGCGCGGACAGCGGAGATCATCTCTGTATACAGCAAG 60

61  AGTCTCTTATACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
|||||
61  AGTCTCTTATACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

61  AGTCTCTTATACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
|||||
61  AGTCTCTTATACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

121  CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
|||||
121  CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

```



```

Best Local Similarity 50.9%; Pred. No. 0.37;
Matches 110; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

427 2555 TGGTTTAAAGAGGATTTAAATTAACCTTAAAGGCTATTCATCTTAATGTTAAAGCAAT 2414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 1446 TGGTTTAAAGAGGATTTAAATTAACCTTAAAGGCTATTCATCTTAATGTTAAAGCAAT 2095
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

429 2415 TTGTATTTAAAGAGGATTTAAATTAACCTTAAAGGCTATTCATCTTAATGTTAAAGCAAT 2474
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 2006 ATGAGCGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2055
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

431 2475 CTAAGTCTGCTAAAGATTAAGAGGATTTAAATTAACCTTAAAGGCTATTCATCTTAATGTTAA 2534
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
432 2006 ATGAGCGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

433 2535 AATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 2570
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 4126 TTATCTTTATTTAAAGGATTTAAATTAACCTTAAAGGCTATTCATCTTAATGTTAAAGCAAT 2161

RESULT 3
US-10-239-676-89
Sequence 89, Application US/10239676
Publication No. US2009082009A1
GENERAL INFORMATION:
APPLICANT: QIAGEN, Alexander
APPLICANT: PLEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
PARENT FILING DATE: 2001-04-24
PRIORITY APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10043529.7
DE 10043826.1
PRIP FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-04-01
NUMAP OF SEQ ID NOS: 228
SEQ ID No 89
LENGTH: 5689
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (3724, 3742, 3755, 3756, 3759, 3760, 3872)
US-10-239-676-89

Query Match 1.4%, Score 45, DB 9; Length 5689;
Best Local Similarity 49.4%; Pred. No. 1.3;
Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

435 2502 TAAATTAATTTCTGCAATATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2361
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 4250 TAAATTAATTAATTAATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 4309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

437 2362 ATGAGCGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2421
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
438 4310 ATGAGCGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4369
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

439 2432 TAAATTAATTTCTGCAATATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 4370 ATGAGCGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4429
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

441 2483 GCTTAAGCTAAGAGGATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 2538
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 4430 ATGAGCGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

100

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:21 11 November 2014

Best Local Similarity 63.2%; Pred. No. 0.39;
Matches 67; Conservative 0; Mismatches 39; Indels 0;
Gaps 0.

Dy
306 GAGTCAATCTACAACTTAAGAATAAATTGTCGGTTTGCGAAGGTGTTATATGTTGCATA <985

Dd
115 GGACCATAGCCAGCATTCTTGGAACCAAGTGGGACTATTAATAAAATATTCGTGTGCCATA 56

```

QY      3066  AATGGAATTAACACCTCCATAATAAAAAAAAAAAAAAAAA 311111
          || | ||| | | | | | | | | | | | | | | | |
Db      55    AATTTAATGTTTCCTGTTAAAAAATAAAAAAAAAAAAA 10

```

RESULT 7
US-09-871-161-408
Sequence 408 Application 116/09871161

```

? Sequence 408 Application HS/64871161
? Publication No. HS20030097665A1
? GENERAL INFORMATION:
? APPLICANT: ENDEGE, MITSON O., ET AL.
? TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
? TITLE OF INVENTION: PRODUCTS- 11
? FILE REFERENCE: CCNDA-260XX
? CURRENT APPLICATION NUMBER: HS/69/871 161
? CURRENT FILING DATE: 2001-05-11
? PRIOR APPLICATION NUMBER: 09/328,111
? PRIOR FILING DATE: 1999-06-08
? PRIOR APPLICATION NUMBER: 66/717,393
? PRIOR FILING DATE: 1999-01-27
? PRIOR APPLICATION NUMBER: 60/098,639
? PRIOR FILING DATE: 1998-08-31
? NUMBER OF SEQ ID NOS: 544
? SOFTWARE: FASTSFY for Windows Version 3.0
? SEQ ID NO: 408
? LENGTH: 588
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc.feature
? LOCATION: (1)..(588)
? OTHER INFORMATION: n = A,T,C or G
? HS-09-871-161-408

```

Query Match	1.48;	Score 43.2;	DB 9;	Length 588;
Best Local Similarity	50.88;	Pred No 0.84;		
Matches 128; Conservative	0;	Mismatches 123;	Indels 1;	Gaps 1.

QY 1667 ATGAGATGTTTATATCAAGAAATGATTTGAGAGAGAGAAATTAACCGGTATGATC 1723
 + ||||| + || + ||||| + || + ||||| +
 Db 104 AGGGGATGTTCTACCAAAAGGGGGTTGGGGGAGAGAGACACACAAAGCTAATAAAA 163

Oy 177 G G A C T A T A C A A A C T G A G A A A T T T C T A T G A A A T G A G A T T A A T A T A T T T T A C G A T T G 178
 Db 164 C C A G A A T T C C C A T T C C C A C A A A A C T G A T G G A A C A A A A T T T T A A G G A T A A A A C A A A A C C C 223

Oy 118 AACGACGACCATTAAGCCACACAAAATGCTATGTATATACTCTTAAGATATA-CATATA 184
 1 11 111111 11 111111 1 11 1111 11 1111
 Db 224 AGTAACTGCTATATTATTAACGCAATATGCTAACCTGTGTTCCTTCTATAGCATCATAT 283

OY 1846 ATCCCTTTAAAGGCAACAGCGATGATTTTACTACAACTAATGATAATGAACAATTGGTAAT] 90
||| ||| | | | | | | | | | | | | | |
Db 284 GTCATGTACTTAGTCTAGTTAAAACATTACATAAGGAACGCGACGAATATATATAAAGAT 343

QY	1906AACATATCCAAA	1917
Db	344 ATATCTCTATA	355

RESULT 8
US-10-239-676-20/c
Sequence 20. Deciphering ms/10219676

```

; PUBLICATIONNO: HSHZ00000000961
;
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

```

1 APPLICANT: BERLIN, Kurt
 2 TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
 3 FILE REFERENCE: 5013.1003
 4 CURRENT APPLICATION NUMBER: 08/10,239, 676
 5 CURRENT FILING DATE: 2002-06-24
 6 PRIOR APPLICATION NUMBER: PCT/EP01/04968
 7

```
DE 10019058.8
DE 10019173.8
DE 10035229.7
DE 10043826.1
PRINT FTTING DATE: 2001-04-06
```

2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228

```

? SEQ ID NO: 20
? LENGTH: 11260
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:

```

Query Match	Score	EB	Length
Best Local Similarity	45.78	Pre	6.02
Query Match	1.48	Score	43.22
		EB	9
		Length	11260

	Matches	150; Conservative	0; Mismatches	178; Indels	0; Gaps
4297 TTGCGCATACCCGCAAGAGCGATTTTGGCTTTATATAGTAAAGAGCAACACATAC					
9861 TTATATCGTAATTTCTTAATATATTATATATATATATATATATATATATATATATAT					

2357 GTTATTATGATTCATTAATTAATCTGAGTCTATTCAGTCAAGTATAAGTAATTT
 9801 TCTACTTTTACACACATTAATAATACACTTAAATTTAAGCTAATCTTTTAAACATTT

[illegible]

Pb 9681 ATACCTTCACGATATTTAAAGATTAAGATTAAGATTAAGATTTACTAAAGATAA
 Qy 2537 TACTATCTATATCAATCAACATCAATGTCTTTCTGTAAGAACTATGATAAATCTAAATCA

Db	9621	AAATCATATTAAATTTTCATATAATTAAACTTAATTAACGCAAAATTACATATC
Oy	2597	TTTTCTTAGCTAAATCATATTACGCA 2524

RESULT 9
US-09-918-995-23852
Sequence 23852, Application US/09918995

```

: Publication No. US20030073623A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Hyseq, Inc.
:
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

```

FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: INS/00/018,005
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: INS/00/235,076
PRIOR FILING DATE: 1999-01-20

```

: NUMBER OF SEQ. ID NOS: 38054
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ. ID NO: 38052
: LENGTH: 478
: TYPE: DNA

```

```

; ORGANISM: Homo sapiens
;
; FEATURE:

```



```

RESULT 13
US-09-938-842A-3170/G
: Sequence 3170, Application US/09-938842A
: Patent No. US20060150378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPT300-3
: CURRENT APPLICATION NUMBER: US/09/938-842A
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/427,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647
: PRIOR FILING DATE: 2001-01-16

```

	Query Match	Best Local Similarity	% (*)	Score	41.8:	DB	10	Length	2804:
	Matches	97:	Conservative	0:	Mismatches	92:	Indels	0:	Gaps
25	1667	AAGGATATTATAAAGAAATTAATCAATTTTGGGTCTGAACCGCATTAACAAGCTTATTCAGC	1726						
140	1586	AGGGGAGGCTTGCCAGCAAAAAGGCGGCCTTTGGGCGAACAGAGAAAGCAAGCAAGAATTAAGAAA	1523						
29	1727	CGACTACAGCAAACTGAGAACTTCTTCTACTGAAATATGAGTTAATATCTCTTTTATATATTC	1786						

Accession	Sequence	Position
D6	2522 CCAGATCCCGCATCCCGCCACAAACATCATGGAGCAAAATTTAAAGCATAAACCAAAACCC	24653
OY	1787 AACAAACGATATACAGACACAAATAGCATATCTTTAAATAAATAGATATATA	18466
D6	2462 ACCAAATACCGATATTACAAACCAATATGTAACCTGTTCCTCTTATATGATATATAT	24033
OY	1847 TCTTTTAA	1855
D6	2492 GTATCTTA	2394

```

1  RESULT: 5
2  US-10-038-001-1
3  : Sequence 1, Application US/10038001
4  : Publication No. US20020187952A1
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: PALMER, Kenneth E.
9  : APPLICANT: POCHE, Gregory P.
10 : APPLICANT: MCCORMICK, Allison
11 :
12 : TITLE OF INVENTION: POLLING CIRCLE REPLY
13 :
14 : TITLE OF INVENTION: VECTORS
15 :
16 : FILE REFERENCE: 008010179C/PUS01
17 :
18 : CURRENT APPLICATION NUMBER: US/10/038,001D
19 :
20 : PRIORITY FILING DATE: 2003-12-20
21 :
22 : PRIOR APPLICATION NUMBER: 09/505,477
23 :
24 : PRIOR FILING DATE: 2000-02-16
25 :
26 : NUMBER OF SEQ ID NOS: 9
27 :
28 : SOFTWARE: FASTSEQ for Windows Version 4.0.
29 :
30 : SEQ ID NO: 1
31 :
32 : LENGTH: 5285
33 :
34 : TYPE: DNA
35 :
36 : ORGANISM: Porcine circovirus
37 :
38 : US-10-038-001-1

```

Query Match	1.38;	Score 41.8;	DB 9;	Length 5285;
Best Local Similarity	75.48;	Pred. No. 8.6;		
Matches 52; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

QY		GGATGCTACTATAACCGCCGCCGTATTGGTGCAGATCACTTTATGTGAACAGGAAAG	60
Dd	293	GCGATCCACACTAATACGGCGCCGCGCATGTGCTGTGGANTTGCGCCTTATTAAATGAGACCGAC	352
QY		AGCTGTTCTCT	69
Dd	353	AGCTGGGTTTT	361

Search completed: July 16, 2003, 00:03:50
Job time : 490 secs

PT beta,beta-carotene 15,15'-dioxygenase protein, nucleic acids and
PT antibodies, useful for production of vitamin A from carotene and gene
therapy of ophthalmological disorders

PS Claim 1; Page 12-13; 37pp; English.

CC Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic
CC conversion of beta-carotene to vitamin A and for raising specific
CC antibodies. Nucleotides encoding all or part of bCOD are useful as
CC primers or probes for specific amplification and/or detection of
CC the gene that encodes bCOD, for isolation of related sequences in
CC other organisms, for determining bCOD levels in humans (to identify
CC subjects requiring vitamin A supplementation) and for detecting
CC mutations in the bCOD gene. The nucleotide encoding bCOD may also
CC be used to transform cells, particularly plant cells, to increase
CC their vitamin A contents (especially in fruits and vegetables) and
CC in gene therapy of subjects who have mutated or deleted forms of the
CC bCOD gene and thus low vitamin A levels and susceptibility to
CC developmental or ophthalmological disorders. Antibodies raised
CC against bCOD are used for detection/quantification of bCOD in
CC immunoassays.

XX Sequence 526 AA;

Query Match 100.0%; Score 526; DR 21; Length 526;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

0Y 1 METFNKKEHPPIKAEGOLPTWLLRGKPMH1IGDTKYNNHMFQGLALHSF 60
0B 1 METFNKKEHPPIKAEGOLPTWLLRGKPMH1IGDTKYNNHMFQGLALHSF 60
0Y 61 TPKNHYVYPSKYI PSITVNCNI EANPIVSEFGTMAVDPCKNIFAKAFSYLSHTPEF 120
0B 61 TPKNHYVYPSKYI PSITVNCNI EANPIVSEFGTMAVDPCKNIFAKAFSYLSHTPEF 120
0Y 121 TDNGLINMKTGDDYATSENFIRKIDPOTLETLKKNYSKVAVANLAISHRHDSACN 180
0B 121 TDNGLINMKTGDDYATSENFIRKIDPOTLETLKKNYSKVAVANLAISHRHDSACN 180
0Y 181 IINNGTSTVDKGRKTYVLFKIPSSVPEKEKKKSCFHLVWVCSIPSRSLQPSYHSECI 240
0B 181 IINNGTSTVDKGRKTYVLFKIPSSVPEKEKKKSCFHLVWVCSIPSRSLQPSYHSECI 240
0Y 241 TENIVTIEOPFKIDYKLAATAYIRGVNMASSCSFHKEDKTMFHFADKTKKKEVSTKFT 300
0B 241 TENIVTIEOPFKIDYKLAATAYIRGVNMASSCSFHKEDKTMFHFADKTKKKEVSTKFT 300
0Y 301 DALVYHNHINAYEEDGHVFDIAYRDNLSYDMFYLLKIDPKPEVNNKLTSPICKRFVY 360
0B 301 DALVYHNHINAYEEDGHVFDIAYRDNLSYDMFYLLKIDPKPEVNNKLTSPICKRFVY 360
0Y 361 PLOYDKDAEYGSNLVLPISATVAVKEKDSITGOREILCEGIELPRVNDYNGKKYKYVY 420
0B 361 PLOYDKDAEYGSNLVLPISATVAVKEKDSITGOREILCEGIELPRVNDYNGKKYKYVY 420
0Y 421 ATEVQMSRPVTKIAKLNVOTKEVLIHMGCEHMCWSEPIFVSPDAREDEDCVLLTCVYSE 480
0B 421 ATEVQMSRPVTKIAKLNVOTKEVLIHMGCEHMCWSEPIFVSPDAREDEDCVLLTCVYSE 480
0Y 481 FNKAPFLLIDAKTFKELGRATVNVEMHLDIOMPIQNMLGAEFE 526
0B 481 FNKAPFLLIDAKTFKELGRATVNVEMHLDIOMPIQNMLGAEFE 526

```

RESULT 2

AAV97314 standard; Protein: 506 AA.

AAV97314:

01-JAN-2001 (first entry)

DE Beta, beta-carotene-15,15'-dioxygenase.

XX Beta, beta-carotene-15,15'-dioxygenase; vitamin A; beta-carotene;
XX transformation; fruit; vegetable; developmental disorder;
XX ophthalmological disorder; antibody; detection; quantification;
XX treatment; therapy.

OS Gallus gallus.

PN EP1031627-A1.

PD 30-AUG-2000.

PF 17-FEB-2000; 2000EP-0102289.

PR 22-FEB-1999; 99EP-0103392.

XX (HOF) HOFFMANN LA KOEHE & CO AG F.

XX Bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;

XX Wp1; 2000-551036/51.

FI Beta,beta-carotene 15,15'-dioxygenase protein, nucleic acids and
PT antibodies, useful for production of vitamin A from carotene and gene
therapy of ophthalmological disorders

XX Disclosure; Page 17-18; 37pp; English.

CC Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic
CC conversion of beta-carotene to vitamin A and for raising specific
CC antibodies. Nucleotides encoding all or part of bCOD are useful as
CC primers or probes for specific amplification and/or detection of
CC the gene that encodes bCOD, for isolation of related sequences in
CC other organisms, for determining bCOD levels in humans (to identify
CC subjects requiring vitamin A supplementation) and for detecting
CC mutations in the bCOD gene. The nucleotide encoding bCOD may also
CC be used to transform cells, particularly plant cells, to increase
CC their vitamin A contents (especially in fruits and vegetables) and
CC in gene therapy of subjects who have mutated or deleted forms of the
CC bCOD gene and thus low vitamin A levels and susceptibility to
CC developmental or ophthalmological disorders. Antibodies raised
CC against bCOD are used for detection/quantification of bCOD in
CC immunoassays.

XX Sequence 506 AA;

Query Match 96.2%; Score 506; DR 21; Length 506;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

0Y 10 EENPEPIKAFVGOGLPTWLLRGKPMH1IGDTKYNNHMFQGLALHSF 69
0B 10 EENPEPIKAFVGOGLPTWLLRGKPMH1IGDTKYNNHMFQGLALHSF 69
0Y 69 PSKYI PSITVNCNI EANPIVSEFGTMAVDPCKNIFAKAFSYLSHTPEF 129
0B 69 PSKYI PSITVNCNI EANPIVSEFGTMAVDPCKNIFAKAFSYLSHTPEF 129
0Y 129 KTGDDYATSENFIRKIDPOTLETLKKNYSKVAVANLAISHRHDSACN 189
0B 129 KTGDDYATSENFIRKIDPOTLETLKKNYSKVAVANLAISHRHDSACN 189
0Y 189 DKGRKTYVLFKIPSSVPEKEKKKSCFHLVWVCSIPSRSLQPSYHSECI 249
0B 189 DKGRKTYVLFKIPSSVPEKEKKKSCFHLVWVCSIPSRSLQPSYHSECI 249
0Y 249 QPKKIDYKLAATAYIRGVNMASSCSFHKEDKTMFHFADKTKKKEVSTKFT 309
0B 249 QPKKIDYKLAATAYIRGVNMASSCSFHKEDKTMFHFADKTKKKEVSTKFT 309
0Y 309 NAYEEDGHVFDIAYRDNLSYDMFYLLKIDPKPEVNNKLTSPICKRFVY 369
0B 309 NAYEEDGHVFDIAYRDNLSYDMFYLLKIDPKPEVNNKLTSPICKRFVY 369

```

```

DB 361 MAFEFHGHVEFIVAPVNSLYDMKYLKLNKEVNNELTSLPTFPFVPLLYOKDAE 360
OY 370 VGSNVLKPLPTSAVAKKKGSIYCPPEILCE3IELFKVNDYNGKKYVAIVQMSPV 429
DB 361 VGSNVLKPLPTSAVAKKKGSIYCPPEILCE3IELFKVNDYNGKKYVAIVQMSPV 420
OY 430 PTKIAKLVNQTKEVLMHGDHGWSEPIFVPSPOAREDBSVLTCVVSSEPNKAPLLI 489
DB 421 PTKIAKLVNQTKEVLMHGDHGWSEPIFVPSPOAREDBSVLTCVVSSEPNKAPLLI 480
OY 490 LDKATFKELGRATVNEHMLDHGMF 515
DB 481 LDKATFKELGRATVNEHMLDHGMF 506

RESULT 3
AAB93108
ID AAB93108 standard; Protein; 547 AA.
AC AAB93108;
XX
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:11972.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001
XX
XX 28-JUN-2000; 2000JP-0116125
XX
XX 26-JUN-1999; 94JP-0244044
XX
XX 27-AUG-1999; 94JP-0300253
XX
XX 11-JAN-2000; 2000JP-0114774
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899
XX
XX (HELI-) HELIX RES INST.
XX
XX Cht T. Isogai T. Nishikawa T. Hayashi K. Saito K. Yamamoto T.
XX Ishii S. Sugiyama T. Makamatsu A. Nagai K. Otsuki T.
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 11972; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification, where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5' end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3' end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3' end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length

```

```

CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and
CC AAB13629 to AAB14742 represent human cDNA sequences; AAB42446 to
CC AAB95893 represent human amino acid sequences; and AAB14629 to AAB17632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 547 AA:
SO
OY Query Match 4.8%; Score 25; DB 22; Length 547;
OY Best Local Similarity 100.0%; Freq No. 2,99e-16;
DB Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 82 NIEANPIVSEFGTMAYPPCKNTP 106
DB 82 NIEANPIVSEFGTMAYPPCKNTP 106

RESULT 4
AAY97313
ID AAY97313 standard; Peptide; 18 AA.
AC AAY97313;
XX
XX
XX 03-JAN-2001 (first entry)
XX
XX Beta, beta-carotene-15,15-dioxygenase tryptic peptide fragment.
XX
XX Beta, beta-carotene-15,15-dioxygenase; vitamin A; beta-carotene;
XX transformation; faulty; vegetable developmental disorder;
XX ophthalmological disorder; antibody; detection; quantification;
XX treatment; therapy.
XX
XX Gallus gallus
XX
XX EP1031627-A1.
XX
XX 03-AUG-2000
XX
XX 17-FEB-2000; 2000EP 0192289.
XX
XX 20-FEB-1999; 94EP-0133082.
XX
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX
XX Reichenau H. Freyberger P. Fritschlin AM. Witzig CW. Wyss A;
XX Wyss M;
XX
XX WPI: 2000-551046/51.
XX
XX Beta, beta-carotene 15,15'-dioxygenase (BCO) is used for enzymatic
XX conversion of beta-carotene to vitamin A and for raising specific
XX antibodies. Nucleotides encoding all or part of bco are useful as
XX primers or probes for specific amplification and/or detection of
XX the gene that encodes bco. For isolation of related sequences in
XX other organisms, for determining bco levels in humans (to identify
XX subjects requiring vitamin A supplementation) and for detecting
XX mutations in the bco gene. The nucleotide encoding bco may also
XX be used to transform cells, particularly plant cells, to increase
XX their vitamin A contents (especially in fruits and vegetables) and
XX in gene therapy of subjects who have mutated or deleted forms of the
XX bco gene and thus low vitamin A levels and susceptibility to
XX developmental or ophthalmological disorders. Antibodies raised
XX against bco are used for detection/quantification of bco in
XX immunoassays.
XX
XX Sequence 18 AA:

```

Query Match 3.4% Score 18; DB 21; Length 18;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 8 NKEHEPEPIKAEEVGGQLP 25
 |||||
 DB 1 NKEHEPEPIKAEEVGGQLP 18

RESULT 5
 AAG62840

ID AAG62840 standard; Protein: 516 AA.

AC AAG62840;

DT 17-SEP-2001 (first entry)

DE Amino acid sequence of a beta-carotene dioxygenase (beta-diox).

KM Beta-carotene dioxygenase; beta-diox; beta-carotene; vitamin A aldehyde;
 transgenic plant.

US Danio rerio.

PN W0200148162-A2.

PD 05-JUL-2001.

PF 22-DEC-2000; 2080WO:EP13144.

PR 24-DEC-1999; 99EP-0125895.

PR 20-MAR-2000; 2000EP-0105822.

PA (GREN-) GRENNOVATION PFLANZENBIOTECHNOLOGIE GMBH.

PI Von Lintig J, Vogt K;

DR WPI: 2001-441713/47.

DR N-PSDB; AAH42170.

PT Dioxygenase (beta-diox) protein that cleaves beta-carotene to form
 vitamin A aldehyde, and polynucleotides encoding them useful for
 producing transgenic bacteria, fungi, plants expressing the polypeptide

PT Claim 7; Page 86-87; 87pp; English.

XX The present sequence represents a beta-carotene dioxygenase (beta-diox)
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic
 CC plants. The transgenic plants have improved nutritional quality or
 CC physiological condition and accumulate vitamin A aldehyde and can take
 CC up beta-carotene from the medium. Expression systems encoding beta-diox
 CC are useful in the study of beta-diox activity. Identification of cDNAs
 CC encoding beta-diox allows the physiological characterization of
 CC mammalian vitamin A metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.

XX Sequence 516 AA;

Query Match 2.5% Score 13; DB 22; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.00037;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 94 GTMAYPDPCKNIF 106
 |||||
 DB 94 GTMAYPDPCKNIF 106

RESULT 6
 ABB71747
 ID ABB71747 standard; Protein: 620 AA.

XX ABB71747;
 AC 26-MAR-2002 (first entry)
 DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 42033.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical).

OS Drosophila melanogaster.

PN W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-0509231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0514150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EM;

DR WPI: 2001-659860/75.

DR N-PSDB; ABL15850.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

PS Disclosure: SEQ ID NO 42033; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL16177), expressed DNA
 CC sequences (ABL16184-ABL16175) and the encoded proteins
 CC (ABR5737-ABR5738).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 620 AA;

Query Match 2.3% Score 11; DB 22; Length 620;
 Best Local Similarity 100.0%; Pred. No. 0.047;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 453 PSEPIFYVSPD 463
 |||||
 DB 550 PSEPIFYVSPD 560

RESULT 7
 AAU04291

ID AAU04291 standard; Protein: 620 AA.

AC AAU04291;

DT 23-OCT-2001 (first entry)

DE Drosophila beta-carotene dioxygenase (Beta-diox 1).

XX Beta-carotene dioxygenase; beta-diox 1; Drosophila; beta-carotene;
 KM lycopene; beta-apocarotenal; beta-ionone; apolycopene; grain seed; corn;
 KM oil seed; palm; chick-pea; diagnostic; therapeutic; ribosome;
 KM retinoid/vitamin A deficiency; beta-diox 1; transgenic; nutrition;
 KM carotene/retinoid pathway; vitamin A aldehyde; retinoid acid; crop.

OS Drosophila melanogaster.
 XX
 PN W0200148163-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 27-DEC-2000; 2000MC-EP13273.
 XX
 PR 24-DEC-1999; 99EP-0125895.
 FF 20-MAR-2000; 2000EP-0105822.
 XX
 PA (GREF-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH
 PI Von Linde J, Vogt K;
 XX
 XX WPI: 2001-425657/45.
 DR N-PSDB: AAS07180.
 XX
 PT Novel isolated mouse, human, *Arabidopsis thaliana* beta-carotene dioxygenase
 (beta-diox II) protein that cleaves beta-carotene and lycopene to yield
 beta-apocarotenal and beta-ionone, and apolycopeneals, respectively.
 XX
 PS Disclosure: Fig 6; 116pp; English.

CC The sequence represents the amino acid sequence of Drosophila beta-
 CC carotene dioxygenase (beta-diox II). Beta-diox II specifically cleaves beta-
 CC carotene and lycopene to form beta-apocarotenal and beta-ionone, and
 CC apolycopeneals, respectively. The DNA is useful for transforming grain
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes;
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or
 CC ribozyme type therapeutic agents and for detecting any abnormality of
 CC endogenous beta-diox II. The beta-diox II specific antisense
 CC oligonucleotides derived from the DNA sequence are useful for dose
 CC response studies in relevant models of retinoid/vitamin A deficiency
 CC during any stage of an organism's development. The nucleic acids are also
 CC useful as probes and as a guideline to define new PCR (polymerase chain
 CC reaction) primers for the cloning of substantially homologous DNA
 CC sequences from other sources. The nucleic acids are also useful for
 CC determining the presence or quantity of beta-diox II nucleic acid and
 CC determining presence and amount of beta-diox II. The polypeptide is also
 CC useful for increasing or decreasing the amount of beta-diox II levels in
 CC a cell or tissue which can modify the level of vitamin A and other
 CC retinoids. Antibodies are useful for studying beta-diox II localization,
 CC screening of an expression library to identify nucleic acids encoding
 CC beta-diox II or the structure of functional domains. The transgenic
 CC lungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
 CC organisms have improved nutritional quality or physiological condition
 CC and accumulate important metabolites of carotene/retinoid pathways such
 CC as vitamin A aldehyde and retinoic acid, beta-carotene or take up beta
 CC carotene from the medium. Expression systems encoding beta-diox II are
 CC useful in the study of beta-diox II activity. Identification of cDNAs
 CC encoding beta-diox I and II allows the physiological characterization of
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.

XX
 XX
 SQ Sequence 620 AA;
 Query Match 2.1%; Score 11; DB 22; Length 620;
 Best Local Similarity 100.0%; Pred No. 0.047;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 453 PSEPIVPSPD 463
 IIIIIIIIIIII
 DB 550 PSEPIVPSPD 560

RESULT 8
 AAG62841
 ID AAG62841 standard; Protein; 620 AA
 XX

AC AAG62841;
 XX
 XX 17-SEP 2001 (first entry)
 DT
 XX
 DE Amino acid sequence of a beta carotene dioxygenase (beta diox).
 XX
 XX Beta-carotene dioxygenase; beta-diox; beta-carotene; vitamin A aldehyde;
 KW transgenic plant.
 XX
 XX Drosophila melanogaster.
 OS
 XX
 PN W0200148162-A2.
 XX
 PD 05-JUL-2001
 XX
 PR 24-DEC-2000; 2000MC-EP13144.
 XX
 PR 24-DEC-1999; 99EP-0125895.
 FF 20-MAR-2000; 2000EP-0105822.
 XX
 PA (GREF-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.
 PI Von Linde J, Vogt K;
 XX
 XX WPI: 2001-441713/47.
 DR N-PSDB: AAB42110.
 XX
 PT D-oxygenase (beta-diox) protein that cleaves beta carotene to form
 PT vitamin A aldehyde, and polynucleotides encoding them useful for
 PT producing transgenic bacteria, fungi, plants expressing the polypeptide
 PT
 PS Claim 7; Fig 6; 87pp; English.

CC The present sequence represents a beta-carotene dioxygenase (beta-diox)
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic
 CC plants. The transgenic plants have improved nutritional quality or
 CC physiological condition and accumulate vitamin A aldehyde and can take
 CC up beta-carotene from the medium. Expression systems encoding beta-diox
 CC are useful in the study of beta-diox activity. Identification of cDNAs
 CC encoding beta-diox allows the physiological characterization of
 CC mammalian vitamin A metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.

XX
 XX
 SQ Sequence 620 AA;
 Query Match 2.1%; Score 11; DB 22; Length 620;
 Best Local Similarity 100.0%; Pred No. 0.047;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 453 PSEPIVPSPD 463
 IIIIIIIIIIII
 DB 550 PSEPIVPSPD 560

RESULT 9
 AAU04293
 ID AAU04293 standard; Protein; 549 AA.
 XX
 AC AAU04293;
 XX
 DT 23-OCT-2001 (first entry)
 DE

DE Zebra-2, beta-carotene dioxygenase (beta-diox II).
 XX
 XX Beta carotene dioxygenase, beta-diox I, zebra-2, beta-carotene;
 KW lycopene; beta-apocarotenal; beta-ionone; apolycopeneal; grain seed; corn;
 KW oil seed; palm; chick-pea; diatom; therapeutic; ribozyme; *Arabidopsis*;
 KW retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;
 KW carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop.
 XX

0S Brachydanio rerio.
 XX
 PN W0200148163-A2.
 XX
 PD 05-JUL-2001.
 XX
 XX 27-DEC-2000: 2000W0-EPI3273
 XX
 XX 24 DEC 1999, 99EP-0125895.
 PR 20 MAR 2000: 2000EP-0105822
 XX
 PA (GREF.) GREENHATTON PLANTENRIJTECHNOLGIE GMBH
 XX
 PI Von Lintig J, Vogt K;
 XX
 DR WPI: 2001-425657/45.
 DR N-PSDB: AAS07195.
 XX
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
 PT (beta-diox II) protein that cleaves beta-carotene and lycopene to yield
 PT beta-apocrotene and beta-ionone, and apolycopene, respectively.
 PS
 PS Claim 2; Fig 14; 116pp; English.
 CC The sequence represents the amino acid sequence of zebra-2, beta-carotene
 CC dioxygenase (beta-diox II). Beta-diox specifically cleaves beta-carotene
 CC and lycopene to form beta-apocrotene and beta-ionone, and
 CC apolycopene, respectively. The DNA is useful for transforming grain
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds, edible seeds e.g. palm;
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes,
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or
 CC ribozyme type therapeutic agents and for detecting any abnormality of
 CC endogenous beta-diox II. The beta-diox II specific antisense
 CC oligonucleotides derived from the DNA sequence are useful for these
 CC response studies in relevant models of retinoid/vitamin A deficiency
 CC during any stage of an organism's development. The nucleic acids are also
 CC useful as probes and as a guideline to define new PCR (polymerase chain
 CC reaction) primers for the cloning of substantially homologous DNA
 CC sequences from other sources. The nucleic acids are also useful for
 CC determining the presence or quantity of beta-diox II nucleic acid and
 CC determining presence and amount of beta-diox II. The polypeptide is also
 CC useful for increasing or decreasing the amount of beta-diox II levels in
 CC a cell or tissue which can modify the level of vitamin A and other
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,
 CC screening of an expression library to identify nucleic acids encoding
 CC beta-diox II or the structure of functional domains. The transgenic
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
 CC organisms have improved nutritional quality or physiological condition
 CC and accumulate important metabolites of carotene/retinoid pathways such
 CC as vitamin A aldehyde and retinoic acid, beta-carotene or take up beta-
 CC carotene from the medium. Expression systems encoding beta-diox II are
 CC useful in the study of beta-diox II activity. Identification of cDNAs
 CC encoding beta-diox I and II allows the physiological characterisation of
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.
 XX
 XX Sequence 549 AA;
 SO
 Query Match 1.98; Score 10; PR 22; Length 549;
 Best Local Similarity 100.0%; Prod No 0.43;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 86 NRIVSSEGT 95
 IIIIIIIIII
 DB 111 NRIVSSEGT 120
 RESULT 10
 AAY97312
 ID AAY97412 standard; peptide: 8 AA.
 XX

AC AAY97312;
 XX
 XX 03 JAN 2001 (first entry)
 XX
 DE Beta, beta-carotene-15,15'-dioxygenase tryptic peptide treatment.
 XX
 XX Beta, beta-carotene-15,15'-dioxygenase, vitamin A beta-carotene;
 XX transformation; fruit; vegetable; developmental disorder;
 XX epithelial; fibroblast; antibody; detection; quantification;
 XX treatment; therapy.
 XX
 OS Gallus gallus.
 XX
 XX EP1031627-A1.
 XX
 XX 30-AUG-2000.
 PD
 XX
 XX 17-FEB-2000; 2000EP-0102289.
 PF
 XX 22-FEB-1999; 99EP-0103482.
 XX
 XX (HOFF.) HOFFMANN LA ROCHE & CO AG F.
 PA
 XX Bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;
 PI Wyss M;
 XX
 DR WPI: 2000-551036/51.
 XX
 PT Beta-carotene-15,15'-dioxygenase protein, nucleic acids and
 PT antibodies, useful for production of vitamin A from carotene and gene
 PT therapy of ophthalmological disorders
 XX
 PS Example 3, Page 7; 37pp; English.
 CC Beta, beta-carotene-15,15'-dioxygenase (bCD) is used for enzymatic
 CC conversion of beta-carotene to vitamin A and for raising specific
 CC antibodies. Nucleotides encoding all or part of bCD are useful as
 CC primers or probes for specific amplification and/or detection of
 CC the gene that encodes bCD, for isolation of related sequences in
 CC other organisms, for determining bCD levels in humans (to identify
 CC subjects requiring vitamin A supplementation) and for detecting
 CC mutations in the bCD gene. The nucleotide encoding bCD may also
 CC be used to transform cells, particularly plant cells, to increase
 CC their vitamin A contents (especially in fruits and vegetables) and
 CC in gene therapy of subjects who have mutated or deleted forms of the
 CC bCD gene and thus low vitamin A levels and susceptibility to
 CC developmental or ophthalmological disorders. Antibodies raised
 CC against bCD are used for detection/quantification of bCD in
 CC immunoassays.
 XX
 XX Sequence 8 AA;
 SO
 Query Match 1.58; Score 8; PR 21; Length 8;
 Best Local Similarity 100.0%; Prod No 7.6005;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 18 AEVGGGP 25
 IIIIIIII
 DB 1 AEVGGGP 8
 RESULT 11
 ABG28616
 ID ABG28616 standard; protein: 46 AA.
 XX
 XX ABG28616;
 XX
 XX 18-FEB-2002 (first entry)
 DT
 XX
 XX Novel human diagnostic protein #28607.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnosis; genetic disorder.
 KW

XX OS Homo sapiens.
 XX PN W0200175067-A2
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001: 2001W0-US00631.
 XX PR 31-MAR-2000: 2000US-0540217
 XX PR 23-AUG-2000: 2000US-0649167
 XX PA (HVSF-) HVSFO INC.
 XX PI Drmanac RT, Liu C, Tang YF;
 XX DR WPI: 2601-63462/73.
 XX DR N-PSDB: AAS92803.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 20: SEQ ID No 58975; 103pp; English.
 XX XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against (I), detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIP0
 CC at ftp.wipo.int/pub/published_pt_sequences.
 XX XX
 SO Sequence 46 AA;
 Query Match 1.5%; Score 8; DB 22; Length 46;
 Best Local Similarity 100.0%; Pred No 4 4;
 Matches 8; Conservation 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 SIPSRL 230
 ID AAY60294 standard; Protein; 82 AA.
 AC AAY60294;
 XX XX
 DT 31-JAN-2000 (first entry)
 XX XX
 DE Human endometrium tumour EST encoded protein 354.
 XX XX
 KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
 KW treatment; uterine; gene therapy; expressed sequence tag
 XX XX
 OS Homo sapiens

XX XX DE19817948-A1.
 XX PN 21-OCT-1999
 XX PD 17-APR-1998: 98DE-1017948.
 XX PR 17-APR-1998: 98DE-1017948.
 XX PR 17-APR-1998: 98DE-1017948.
 XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 XX DR WPI: 1999-591957/51.
 XX DR N-PSDB: AA242095.
 PT New nucleic acid sequences expressed in uterine cancer tissues, and
 PT derived polypeptides, for treatment of uterine and endometrial cancer
 PT and identification of therapeutic agents.
 PS Claim 23: Page 416; 444pp; German.
 XX XX This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for
 CC generation of specific antibodies. (A) are identified by assembling ESTs
 CC (expressed sequence tags) from a particular tissue type before comparison
 CC of expression patterns. This allows a significantly longer fragment of
 CC the gene to be revealed, so should reduce the number of failures
 CC associated with the fact that ESTs from different libraries may represent
 CC different parts of the same unknown gene, distorting the estimated
 CC frequency of occurrence in a particular tissue. AAY5941 Y60328 represent
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in AAC41981 C42101.
 XX XX
 SO Sequence 82 AA;
 Query Match 1.5%; Score 8; DB 20; Length 82;
 Best Local Similarity 100.0%; Pred No. 7 6;
 Matches 8; Conservation 0; Mismatches 0; Indels 0; Gaps 0;
 QY 21 QGOLPTWL 28
 ID AAB73956 standard; Protein; 85 AA.
 AC AAB73956;
 XX XX
 DT 29 MAY-2001 (first entry)
 XX XX
 DE Canine mutant RPE65.
 XX XX
 KW Canine, dog, RPE65, canine retinal pigment epithelium 65, RPE65;
 KW congenital stationary night blindness; CSNB; mutation detection;
 KW mutant.
 XX XX
 OS Canis familiaris.
 XX PN US6201114-R1.
 XX PD 13-MAR-2001.
 XX PR 30 AUG-1999: 98US 0085259.
 XX PR 06-OCT-1998: 98US-0103219.

XX BA (CGRK) CÖRNEIL RES FOUND INC.
XX XX
X1 Aguirre GD, Arland GW, Ray K:
XX WPJ: 2001-265168/27.
XX DR
XX PT Novel nucleic acid molecule encoding canine retinal pigment epithelium
PT 65, where presence of mutation in one or both alleles is indicative of
PT a carrier of, or dog affected with congenital stationary night
PT blindness -
XX PS
PS Disclosure; Fig 1C; 25pp; English.
XX CC
CC The present sequence is a mutated version of canine retinal pigment
CC epithelium 65 (PP65). The gene encoding this protein contains an AACA
CC deletion at nucleotides 487-490 and is responsible for the disease
CC congenital stationary night blindness (CSNB). The gene is useful for
CC identifying dogs which are genetically normal, or are carriers of, or
CC affected with congenital stationary night blindness (CSNB) This allows
CC a breeder to eliminate the carrier from the breeding stock of to breed
CC carriers with genetically normal dogs.
XX SQ
SQ Sequence: 205 AA:

Query Match 1.5%, Score 8, E# 22; Length 205;
Host Local Similarity 100.0%; Pred.No.18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 99 PDCKKNIF 106
| | | | | | | |
DB 109 PDCKKNIF 116

RESULT 14
AAG41938
ID AAG41938 standard; Protein: 318 AA.
XX AC
AC AAG41938:
XX DT
DT 18-OCT-2000 (first entry)
XX DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52240.
XX KW
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS
OS Arabidopsis thaliana.
XX FN
FN EP1031405-A2.
XX PD
PD 06-SEP-2000.
XX XX
XX 25 FEB 2000, 2060BP 0301439
XX XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR	04-MAY-1999	9905-0132485
PR	05-MAY-1999	9905-0132485
PR	06-MAY-1999	9905-0132485
PR	07-MAY-1999	9905-0132485
PR	11-MAY-1999	9905-0132485
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134221
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134478
PR	18-MAY-1999	9905-0134491
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135353
PR	24-MAY-1999	9905-0135659
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136362
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137252
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137505
PR	06-JUN-1999	9905-0137724
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139465
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139617
PR	23-JUN-1999	9905-0140053
PR	24-JUN-1999	9905-0140054
PR	24-JUN-1999	9905-0140055
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142156
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142430
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0143097
PR	12-JUL-1999	9905-0143297
PR	13-JUL-1999	9905-0143342
PR	14-JUL-1999	9905-0143524
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145088
PR	21-JUL-1999	9905-0145088

PR 22-JUL-1999: 99US-0145085
 PR 22-JUL-1999: 99US-0145087
 PR 22-JUL-1999: 99US-0145089
 PR 22-JUL-1999: 99US-0145192
 PR 23-JUL-1999: 99US-0145145
 PR 23-JUL-1999: 99US-0145218
 PR 23-JUL-1999: 99US-0145224
 PR 26-JUL-1999: 99US-0145276
 PR 27-JUL-1999: 99US-0145913
 PR 27-JUL-1999: 99US-0145919
 PR 28-JUL-1999: 99US-0145951
 PR 02-AUG-1999: 99US-0146586
 PR 02-AUG-1999: 99US-0146588
 PR 02-AUG-1999: 99US-0146589
 PR 03-AUG-1999: 99US-0147038
 PR 04-AUG-1999: 99US-0147204
 PR 05-AUG-1999: 99US-0147302
 PR 05-AUG-1999: 99US-0147192
 PR 06-AUG-1999: 99US-0147303
 PR 06-AUG-1999: 99US-0147416
 PR 09-AUG-1999: 99US-0147493
 PR 09-AUG-1999: 99US-0147935
 PR 10-AUG-1999: 99US-0148171
 PR 11-AUG-1999: 99US-0148319
 PR 12-AUG-1999: 99US-0148341
 PR 13-AUG-1999: 99US-0148565
 PR 13-AUG-1999: 99US-0148684
 PR 16-AUG-1999: 99US-0149488
 PR 17-AUG-1999: 99US-0149175
 PR 18-AUG-1999: 99US-0149426
 PR 20-AUG-1999: 99US-0149722
 PR 20-AUG-1999: 99US-0149723
 PR 20-AUG-1999: 99US-0149929
 PR 23-AUG-1999: 99US-0149902
 PR 23-AUG-1999: 99US-0149930
 PR 25-AUG-1999: 99US-0150566
 PR 26-AUG-1999: 99US-0150884
 PR 27-AUG-1999: 99US-0151065
 PR 27-AUG-1999: 99US-0151066
 PR 27-AUG-1999: 99US-0151080
 PR 30-AUG-1999: 99US-0151303
 PR 31-AUG-1999: 99US-0151338
 PR 01-SEP-1999: 99US-0151930
 PR 07-SEP-1999: 99US-0152363
 PR 10-SEP-1999: 99US-0153070
 PR 13-SEP-1999: 99US-0154758
 PR 15-SEP-1999: 99US-0154018
 PR 16-SEP-1999: 99US-0154039
 PR 20-SEP-1999: 99US-0154779
 PR 22-SEP-1999: 99US-0155136
 PR 23-SEP-1999: 99US-0155486
 PR 24-SEP-1999: 99US-0155659
 PR 28-SEP-1999: 99US-0156458
 PR 28-SEP-1999: 99US-0156466
 PR 04-OCT-1999: 99US-0157117
 PR 05-OCT-1999: 99US-0157753
 PR 06-OCT-1999: 99US-0157865
 PR 07-OCT-1999: 99US-0158029
 PR 08-OCT-1999: 99US-0158422
 PR 12-OCT-1999: 99US-0158369
 PR 13-OCT-1999: 99US-0158293
 PR 13-OCT-1999: 99US-0158294
 PR 13-OCT-1999: 99US-0158295
 PR 14-OCT-1999: 99US-0159329
 PR 14-OCT-1999: 99US-0159438
 PR 14-OCT-1999: 99US-0159437
 PR 14-OCT-1999: 99US-0159438
 PR 18-OCT-1999: 99US-0159684
 PR 21-OCT-1999: 99US-0160741
 PR 21-OCT-1999: 99US-0160767

PR 21-OCT-1999: 99US-0160768
 PR 21-OCT-1999: 99US-0160770
 PR 21-OCT-1999: 99US-0160814
 PR 21-OCT-1999: 99US-0160815
 PR 21-OCT-1999: 99US-0160980
 PR 21-OCT-1999: 99US-0160981
 PR 21-OCT-1999: 99US-0160989
 PR 25-OCT-1999: 99US-0161404
 PR 25-OCT-1999: 99US-0161405
 PR 25-OCT-1999: 99US-0161406
 PR 25-OCT-1999: 99US-0161359
 PR 25-OCT-1999: 99US-0161360
 PR 25-OCT-1999: 99US-0161661
 PR 25-OCT-1999: 99US-0161920
 PR 28-OCT-1999: 99US-0161993
 PR 29-OCT-1999: 99US-0162142

Query Match 1.5% Score 8: DB 21: Length 318:
 Best Local Similarity 100.0% Ident. No. 27:
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Caps 0:

Oy 491 DAKFKEL 498
 Db 72 DAKFKEL 79

RESULT 15
 AAC41937 standard: Protein: 372 AA.
 XX
 AC AAC41937:
 XX
 XX 18-OCT-2000 (first entry)
 XX
 XX Arabidopsis thaliana protein fragment Sbj ID NO: 52239.
 DE
 XX
 XX Protein identification, signal transduction pathway, metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX
 XX Arabidopsis thaliana.
 OS
 XX
 XX FPI033405-A2.
 FN
 XX
 XX 06-SEP-2000.
 FD
 XX
 XX 25-SEP-2000: 2000EP-0301439
 XX
 XX 25-SEP-1999: 99US-0131825
 PR 05-MAY-1999: 99US-0132180
 PR 05-MAY-1999: 99US-0132448
 PR 05-MAY-1999: 99US-0132588
 PR 26-MAY-1999: 99US-0132604
 PR 26-MAY-1999: 99US-0132685
 PR 01-APR-1999: 99US-0132742
 PR 05-APR-1999: 99US-0132834
 PR 05-APR-1999: 99US-0128714
 PR 16-APR-1999: 99US-0129845
 PR 19-APR-1999: 99US-0130077
 PR 21-APR-1999: 99US-0130449
 PR 21-APR-1999: 99US-0130510
 PR 21-APR-1999: 99US-0130891
 PR 28-APR-1999: 99US-0131449
 PR 28-APR-1999: 99US-0132448
 PR 30-APR-1999: 99US-0132407
 PR 04-MAY-1999: 99US-0132484
 PR 05-MAY-1999: 99US-0132485
 PR 05-MAY-1999: 99US-0132486
 PR 05-MAY-1999: 99US-0132487
 PR 07-MAY-1999: 99US-0132863
 PR 11-MAY-1999: 99US-0134256
 PR 14-MAY-1999: 99US-0134218

PR 14-MAY-1999; 990S-0134219
PR 14-MAY-1999; 990S-0134221
PR 14-MAY-1999; 990S-0134270
PR 18-MAY-1999; 990S-0134378
PR 19-MAY-1999; 990S-0134941
PR 20-MAY-1999; 990S-0135124
PR 21-MAY-1999; 990S-0135353
PR 24-MAY-1999; 990S-0135629
PR 25-MAY-1999; 990S-0136021
PR 27-MAY-1999; 990S-0136392
PR 28-MAY-1999; 990S-0136782
PR 01-JUN-1999; 990S-0137222
PR 03-JUN-1999; 990S-0137528
PR 04-JUN-1999; 990S-0137502
PR 07-JUN-1999; 990S-0137724
PR 08-JUN-1999; 990S-0138094
PR 10-JUN-1999; 990S-0138540
PR 10-JUN-1999; 990S-0138847
PR 14-JUN-1999; 990S-0139119
PR 16-JUN-1999; 990S-0139452
PR 16-JUN-1999; 990S-0139453
PR 17-JUN-1999; 990S-0139492
PR 18-JUN-1999; 990S-0139454
PR 18-JUN-1999; 990S-0139455
PR 18-JUN-1999; 990S-0139456
PR 18-JUN-1999; 990S-0139457
PR 18-JUN-1999; 990S-0139458
PR 18-JUN-1999; 990S-0139459
PR 18-JUN-1999; 990S-0139460
PR 18-JUN-1999; 990S-0139461
PR 18-JUN-1999; 990S-0139462
PR 18-JUN-1999; 990S-0139463
PR 18-JUN-1999; 990S-0139750
PR 18-JUN-1999; 990S-0139763
PR 21-JUN-1999; 990S-0139817
PR 22-JUN-1999; 990S-0139899
PR 23-JUN-1999; 990S-0140353
PR 23-JUN-1999; 990S-0140354
PR 24-JUN-1999; 990S-0140645
PR 28-JUN-1999; 990S-0140823
PR 29-JUN-1999; 990S-0140891
PR 30-JUN-1999; 990S-0141287
PR 01-JUL-1999; 990S-0141842
PR 01-JUL-1999; 990S-0142154
PR 02-JUL-1999; 990S-0142065
PR 06-JUL-1999; 990S-0142390
PR 08-JUL-1999; 990S-0142803
PR 12-JUL-1999; 990S-0142977
PR 13-JUL-1999; 990S-0143542
PR 14-JUL-1999; 990S-0143624
PR 15-JUL-1999; 990S-0144005
PR 16-JUL-1999; 990S-0144085
PR 16-JUL-1999; 990S-0144086
PR 19-JUL-1999; 990S-0144325
PR 19-JUL-1999; 990S-0144331
PR 19-JUL-1999; 990S-0144372
PR 19-JUL-1999; 990S-0144333
PR 19-JUL-1999; 990S-0144334
PR 19-JUL-1999; 990S-0144335
PR 20-JUL-1999; 990S-0144352
PR 20-JUL-1999; 990S-0144632
PR 20-JUL-1999; 990S-0144884
PR 21-JUL-1999; 990S-0144814
PR 21-JUL-1999; 990S-0145086
PR 21-JUL-1999; 990S-0145088
PR 22-JUL-1999; 990S-0145085
PR 22-JUL-1999; 990S-0145087
PR 22-JUL-1999; 990S-0145089
PR 22-JUL-1999; 990S-0145192
PR 23-JUL-1999; 990S-0145218
PR 23-JUL-1999; 990S-0145219
PR 23-JUL-1999; 990S-0145224

PR 26-JUL-1999; 990S-0145976
PR 27-JUL-1999; 990S-0145973
PR 27-JUL-1999; 990S-0145974
PR 27-JUL-1999; 990S-0145975
PR 28-JUL-1999; 990S-0145951
PR 02-AUG-1999; 990S-0146386
PR 02-AUG-1999; 990S-0146388
PR 02-AUG-1999; 990S-0146389
PR 03-AUG-1999; 990S-0147038
PR 04-AUG-1999; 990S-0147204
PR 04-AUG-1999; 990S-0147202
PR 05-AUG-1999; 990S-0147126
PR 05-AUG-1999; 990S-0147260
PR 06-AUG-1999; 990S-0147303
PR 06-AUG-1999; 990S-0147416
PR 06-AUG-1999; 990S-0147443
PR 09-AUG-1999; 990S-0147935
PR 10-AUG-1999; 990S-0148171
PR 11-AUG-1999; 990S-0148319
PR 12-AUG-1999; 990S-0148341
PR 13-AUG-1999; 990S-0148665
PR 13-AUG-1999; 990S-0148684
PR 16-AUG-1999; 990S-0149368
PR 17-AUG-1999; 990S-01494175
PR 18-AUG-1999; 990S-0149426
PR 20-AUG-1999; 990S-0149722
PR 20-AUG-1999; 990S-0149723
PR 20-AUG-1999; 990S-0149929
PR 23-AUG-1999; 990S-0149902
PR 23-AUG-1999; 990S-0149930
PR 25-AUG-1999; 990S-0150566
PR 25-AUG-1999; 990S-0150844
PR 27-AUG-1999; 990S-0151065
PR 27-AUG-1999; 990S-0151066
PR 27-AUG-1999; 990S-0151080
PR 30-AUG-1999; 990S-0151303
PR 31-AUG-1999; 990S-0151438
PR 01-SEP-1999; 990S-0151930
PR 07-SEP-1999; 990S-0152363
PR 10-SEP-1999; 990S-0153070
PR 13-SEP-1999; 990S-0153758
PR 15-SEP-1999; 990S-0154018
PR 16-SEP-1999; 990S-0154039
PR 20-SEP-1999; 990S-0154779
PR 22-SEP-1999; 990S-0155139
PR 23-SEP-1999; 990S-0155486
PR 24-SEP-1999; 990S-0155458
PR 24-SEP-1999; 990S-0156566
PR 29-SEP-1999; 990S-0157117
PR 04-OCT-1999; 990S-0157753
PR 05-OCT-1999; 990S-0157865
PR 06-OCT-1999; 990S-0158029
PR 07-OCT-1999; 990S-0158232
PR 08-OCT-1999; 990S-0158269
PR 12-OCT-1999; 990S-0159295
PR 13-OCT-1999; 990S-0159294
PR 13-OCT-1999; 990S-0159329
PR 14-OCT-1999; 990S-0159330
PR 14-OCT-1999; 990S-0159331
PR 14-OCT-1999; 990S-0159637
PR 14-OCT-1999; 990S-0159638
PR 18-OCT-1999; 990S-0159584
PR 21-OCT-1999; 990S-0160741
PR 21-OCT-1999; 990S-0160767
PR 21-OCT-1999; 990S-0160768
PR 21-OCT-1999; 990S-0160770
PR 21-OCT-1999; 990S-0160814
PR 21-OCT-1999; 990S-0160815
PR 22-OCT-1999; 990S-0160980
PR 22-OCT-1999; 990S-0160981
PR 22-OCT-1999; 990S-0160989

PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 1.5%; Score 8; DB 21; Length 372;

Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 DAKTFKEL 498
|||||||
DB 126 DAKTFKEL 133

Search completed: July 15, 2003, 09:40:39
Job time : 74 secs

GenCore version 5.1.4
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2003, 09:40:45 : Search time 43 seconds

(without alignments)
1175 970 Million cell updates/sec

Title: US-10-053-192-1

Sequence: 1 METFNKMKKEHPIKAEV MLLLLHGMFLPNDLSAEITE 525

Scoring table: 0:100
Gapop 60.0, Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR-71.*

1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.5	533	A47143	retinal pigment mi
2	8	1.5	580	F81042	hemolysin activati
3	8	1.5	580	A81989	probable periplasm
4	8	1.5	856	I58411	protein-tyrosine k
5	8	1.5	876	I49152	protein-tyrosine k
6	8	1.5	880	JC4166	protein-tyrosine k
7	8	1.5	880	RS3743	protein-tyrosine k
8	7	1.3	78	A97719	hypothetical prote
9	7	1.3	111	B72482	hypothetical prote
10	7	1.3	124	T05467	hypothetical prote
11	7	1.3	125	B87322	chemotaxis protein
12	7	1.3	140	H71197	hypothetical prote
13	7	1.3	177	T35580	hypothetical prote
14	7	1.3	180	G69222	probable acetyltra
15	7	1.3	181	T49476	molybdenum formylm
16	7	1.3	183	A13566	related to 50S RIB
17	7	1.3	187	F95080	hypothetical prote
18	7	1.3	193	T17952	conserved domain p
19	7	1.3	216	C79843	hypothetical prote
20	7	1.3	233	T40775	serine esterase ho
21	7	1.3	243	SE6405	ribosome biosynthe
22	7	1.3	243	P91273	hypothetical 26.6K
23	7	1.3	243	U66114	hypothetical prote
24	7	1.3	243	A11049	hypothetical prote
25	7	1.3	258	E90126	probable tRNA/tRNA
26	7	1.3	260	T16877	hypothetical prote
27	7	1.3	261	T43579	hypothetical prote
28	7	1.3	264	A64637	type III secretion
29	7	1.3	267	T43313	hypothetical prote
					Radl like protein

30	7	1.3	274	B71877	hypothetical prote
31	7	1.3	282	G97947	hypothetical prote
32	7	1.3	290	A11730	lactobacillus delb
33	7	1.3	297	T19888	hypothetical prote
34	7	1.3	313	F83469	hypothetical prote
35	7	1.3	344	D22735	hypothetical prote
36	7	1.3	347	G66741	hypothetical nox3
37	7	1.3	352	A81055	unknown protein F1
38	7	1.3	365	B69793	tRNA (tracil-5-)-m
39	7	1.3	467	A83582	hypothetical prote
40	7	1.3	378	JC5658	sarcosine oxidase
41	7	1.3	379	E93469	LIM domain contain
42	7	1.3	380	C93469	hypothetical prote
43	7	1.3	382	G64509	conserved hypotet
44	7	1.3	382	A97892	gap junction prote
45	7	1.3	384	T38544	probable exoglyph

ALIGNMENTS

RESULT 1

A47143

retinal pigment, microsome protein, PPE5, epithelial-specific, bovine

N:Alternate names: membrane receptor p63; retinol-binding protein receptor

C:Species: Bos primigenius taurus (cattle)

C>Date: 03-May-1994 #sequence, revision 03-May-1994 #text, change 05-Nov-1999

C:Accession: A47143; A48017; S28503

J:Hamel, C P.; Tsiilon, E.; Pfeiffer, B A.; Hooks, J J.; Detrick, R.; Redmond, T M

R: Biol. Chem. 268, 15751-15757, 1993

A:Title: Molecular cloning and expression of PPE5, a novel retinal pigment epithelium

A:Reference number: A47143, MIM:633401B1, EMBL:034040

A:Accession: A47143

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-533 :HAM.

A:Cross-references: GB L11356, NIE-9163656; PIRN:AA37306.1; PIR:q163657

F:Bank. C.O.; Levy, F.; Hellman, U.; Weststedt, C.; Eriksson, U.

J: Biol. Chem. 268, 20540-20546, 1993

A:Title: The retinal pigment epithelial membrane receptor for plasma retinol-binding

A:Reference number: A48017; MIM:9338633; PMID:8397208

A:Accession: A48017

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-340 :T, 342-533 :BAC.

A:Cross-references: FMR, X66277; NITRG63, T1104 CAA4668.1; E16: 9564

C:Keywords: membrane protein

Query Match

Best local similarity 100.0%, Pred. No. 14

Methods: 8: Consistency, 0: Mismatches, 0: Indels, 0: Gaps, 0:

Qy 93 PDPCKNIF 106

Db 109 PDPCKNIF 116

RESULT 2

F81042

hemolysin activation protein Heeb, probable NMB1780 [imported] Neisseria meningitidis

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence, revision 31-Mar-2000 #text, change 19-Jan-2001

C:Accession: F81042

F: Tettelin, H.; Sandberg, N. I.; Hultberg, S. J.; Tettelin, A. G.; Nelson, K. E.; Eisen, R. I.; H. G.; Van der Vliet, A. J.; Smith, H. G.; Fraser, C. M.; McLeod, E. R.; Cappuccin, R. J.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sim, L.; Smith, H. G.; Fraser, C. M.; McLeod, E. R.; Cappuccin, R. J.

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain M58.

A:Reference number: A81000, MIM:20175755; PMID:10710307

A:Accession: F81042

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-580 <TET>
 A:Cross-references: GR:AF002528; GR:AF002598; NID:47227334; PIRN:AA42120.1; PIR:4722703
 A:Experimental source: serogroup A, strain MCS8
 C:Genetics:

Query Match 1.5% Score 8; PR 2; Length 580;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 407 VNYDNGK 414
 |||||
 DB 325 VNYDNGK 332

RESULT 3

AB1989
 probable periplasmic protein NMA0687 [imported] - Neisseria meningitidis (strain 22491)
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: AB1989
 P:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
 J.; Holtrop, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 N.; et al. 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
 A:Reference number: AB1989; MIM:2032556; PMID:10761919
 A:Accession: AB1989
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-580 <PAR>
 A:Cross-references: GR:AL12773; GR:AL157959; NID:47379120; PIRN:CA681973.1; PIR:47379120
 A:Experimental source: serogroup A, strain 22491
 C:Genetics:

Query Match 1.5% Score 8; DB 2; Length 580;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 407 VNYDNGK 414
 |||||
 DB 325 VNYDNGK 332

RESULT 4

158411
 protein-tyrosine kinase (EC 2.7.1.112) b7 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 04-Feb-2000
 C:Accession: 158411
 R:Fujimoto, J.; Yamamoto, T.
 Oncogene 9, 693-698, 1994
 A:Title: b7, a mouse gene encoding a novel receptor-type protein-tyrosine kinase, is put
 A:Reference number: 158411; MIM:194150990; PMID:8108111
 A:Accession: 158411
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-856 <RKS>
 A:Cross-references: GR:017493; NID:4801867; PIRN:AA04216.1; PIR:4801867
 C:Superfamily: protein-tyrosine kinase axl; fibronectin type III repeat homology; immunoglobulin domain; phosphotransferase; tyrosine-specific protein kinase
 C:Keywords: ATP, glycoprotein, phosphotransferase, tyrosine-specific protein kinase
 F:142-191/Domain: immunoglobulin homology <IMM>
 F:501-778/Domain: protein kinase homology <KIN>
 F:509-517/Region: protein kinase ATP-binding motif

Query Match 1.5% Score 8; DB 2; Length 856;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 52 DGIALLHS 59
 |||||
 DB 236 DGIALLHS 243

RESULT 5
 149152
 protein-tyrosine kinase (EC 2.7.1.112) tyro3, isoform B - mouse

N:Alternate names: tyrosine kinase growth factor receptor
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Feb-2000
 C:Accession: 149152; EMBL:149152

R:Schulz, N.T.; Pauliac, C.L.; Lee, L.; Zhou, K.
 Brain Res. Mol. Brain Res. 28, 271-280, 1995

A:Title: Isolation and expression analysis of tyro3, a murine growth factor receptor
 A:Reference number: 149151; MIM:95240399; PMID:7723626
 A:Accession: 149152

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-876 <RSS>

A:Cross-references: EMBL:U08343; NID:4687627; PIRN:AA26443.1; PIR:4687628
 R:Riesewer, L.G.; Giamola, D.M.; Emerson, S.G.
 Oncogene 10, 2239-2242, 1995

A:Title: Identification of alternative exons, including a novel exon, in the tyrosine
 A:Reference number: 148861; MIM:9530487; PMID:7784069
 A:Accession: 148861

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-88 <RE2>

A:Cross-references: EMBL:U08343; NID:4687627; PIRN:AA26443.1; PIR:4687628
 C:Genetics:

A:Gene: tyro3; Etk2/tyro3
 A:Introns: 28/1

C:Superfamily: protein-tyrosine kinase axl; fibronectin type III repeat homology; immunoglobulin domain; phosphotransferase; tyrosine kinase
 C:Keywords: ATP, glycoprotein, growth factor receptor, phosphotransferase, tyrosine kinase
 F:43-105/Domain: immunoglobulin homology <IMM>
 F:308-331/Domain: fibronectin type III repeat homology <3FR>
 F:502-773/Domain: protein kinase homology <KIN>
 F:510-518/Region: protein kinase ATP-binding motif

Query Match 1.5% Score 8; DB 2; Length 876;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 52 DGIALLHS 59
 |||||
 DB 236 DGIALLHS 243

RESULT 6

JC4166
 protein-tyrosine kinase (EC 2.7.1.112) tyro3 precursor - rat

N:Alternate names: protein-tyrosine kinase src; receptor-type tyrosine kinase lse
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: JC4166; PIR:185

R:Ohashi, K.; Hudaib, S.; Ichikawa, Y.; Nakamura, I.; Mizuno, K.
 J. Biochem. 117, 1267-1275, 1995
 A:Title: Molecular cloning and in situ localization in the brain of rat Src receptor
 A:Reference number: JC4166; MIM:46104999; PMID:7490270

A:Accession: JC4166
 A:Molecule type: mRNA

A:Cross-references: PIR:185; NID:4148195; PIRN:AA07111.1; PIR:4148195
 A:Experimental source: brain
 A>Note: It is uncertain whether Met-1 or Met-7 is the initiator

R:Li, C.; Lemke, G.
 Neuron 6, 691-704, 1991

A:Title: An extended family of protein tyrosine kinase genes differentially expressed
 A:Reference number: PIR:185; MIM:4122566; PMID:2055425
 A:Accession: PIR:185

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 650-703 <LA1>
 A:Experimental source: sciatic nerve
 C:Comment: This receptor plays an important role in development, function, and maintenance of tyro-3

hypothetical protein APE2498 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #ext_change 20-Jun-2000
 C:Accession: H72482
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jinno, K.; Takahawa, H.; Takamizawa, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, S.; K DNA Res. 6, 84-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic ctenarchaeon, Aeropyrum pernix strain K1
 A:Reference number: A72450; MIMD:99310339; PMID:10382966
 A:Accession: H72482
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-111 <KAW>
 A:Cross-references: DDBJ:AB000064; NID:95105945; P10N:HA01514.1; PID:95106203
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE2498
 C:Superfamily: Aeropyrum pernix hypothetical protein APE2498

Query Match 1.38; Score 7; DB 2; Length 111,
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

OY 165 AVNLATS 171
 DB 35 AVNLATS 41

RESULT 10
 T05467
 hypothetical protein T805.60 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 20-Apr-2000
 C:Accession: T05467
 R:Ravay, M.; Wiedler, H.; Wambutt, P.; Hancock, J.; Mewes, H.W.; Mayor, K.F.X.; Schmolze submitted to the Protein Sequence Database, February 1998
 A:Reference number: Z15417
 A:Accession: T05467
 A:Molecule type: DNA
 A:Residues: 1-124 <EBV>
 A:Cross-references: EMBL:AL021890
 A:Experimental source: cultivar Columbia; BAC clone T805
 C:Genetics:
 A:Map position: 4
 A:Introns: 40/2; 92/3
 A:Note: T805.60
 C:Superfamily: hypothetical protein YC1033c

Query Match 1.38; Score 7; DB 2; Length 124,
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 178 PTSATAV 384
 DB 2 PTSATAV 8

RESULT 11
 B87322
 chemotaxis protein CheY [Imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #ext_change 10-May-2001
 C:Accession: B87322
 R:Kierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Leub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapell, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete genome sequence of Caulobacter crescentus.
 A:Reference number: A87249; MIMD:21173698; PMID:11256647
 A:Accession: B87322
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-125 <STO>

A:Cross-references: CB:AE005673, NID:913421788, P10N:AAK22574.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0588
 C:Superfamily: chemotaxis cheY protein, response regulator homolog

Query Match 1.38; Score 7; DB 2; Length 125,
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 52 DGLALLH 58
 DB 61 DGLALLH 67

RESULT 12
 H71197
 hypothetical protein PH1854 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #ext_change 28-Jul-2000
 C:Accession: H71197
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Se M.; Ohjima, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, S.; Kudo, N.; Ogi DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A:Reference number: A71000; MIMD:98344137; PMID:9679194
 A:Accession: H71197
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-140 <KAW>
 A:Cross-references: CB:AP000007; NID:94736134; P10N:HA010975.1; PID:94736292
 A:Experimental source: strain OT3
 A:Note: This accession replaces an interim accession for a sequence replaced by Genba C:Genetics:
 A:Gene: PH1854
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1854

Query Match 1.38; Score 7; DB 2; Length 140;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 347 NKLTSP 353
 DB 13 NKLTSP 19

RESULT 13
 T35580
 probable acetyltransferase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #ext_change 21-Jul-2000
 C:Accession: T35580
 R:Saunders, D.; Harris, D.; Parkhill, J.; Barrett, B.G.; Kalandraam, M.A. submitted to the EMBL Data Library, August 1998
 A:Reference number: Z21583
 A:Accession: T35580
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-177 <SAU>
 A:Cross-references: EMBL:AF031317, P10N:CAA20407.1; GSPDB:GN00070, SGPDB:SG0064.29
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SGPDB:SG0064.29
 C:Superfamily: Escherichia coli ribosomal protein-alanine N-acetyltransferase flm

Query Match 1.38; Score 7; DB 2; Length 177;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 313 EEDGHV 319
 DB 63 EEDGHV 69

RESULT 14

G69222

molybdenum formylmethanofuran dehydrogenase, subunit E Methanobacterium thermoautotroph

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Oct-1999

C:Accession: G69222

R:Smith, D.R.; Donner, S.; Stamm, L.A.; Donoherty, C.; Lee, H.; Daniels, J.; Althage, T.;
Qiu, D.; Spadafino, K.; Vaites, R.; Wang, Y.; Wierzbowski, J.; Gilson, P.; Jiwani, N.
K.; S. Church, G.M.; Daniels, C.J.; Mo, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997

J:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func

A:Reference number: A63000, MIMD:98037514, PMID:9371463

A:Accession: G69222

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-180 <MTH>

A:Cross-references: GB:AE000867; GB:AE000868; NID:9522009; P10N:AA09445.1; P1D:9264201

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH917

A:Start codon: GTG

C:Superfamily: Methanosarcina barkeri molybdenum formylmethanofuran dehydrogenase chain

Query Match

1.3%; Score 7; DB 2; Length 180;

Best local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 VPEKEK 211

Db 147 VPEKEK 153

RESULT 15

T49476

related to 50S RIBOSOMAL PROTEIN L2 [imported] - Neurospora crassa

N:Alternate names: protein B14D6.280

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49476

R:Schulze, H.; Altmann, V.; Hehlsel, J.; Brandt, P.; Farlmann, R.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: 225022

A:Accession: T49476

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-181 <SCH>

A:Cross-references: EMBL AL356173; GSPR: G69016; NCSP: B14D6.280

A:Experimental source: BAC clone B14D6; strain OK74A

C:Genetics:

A:Gene: NCSP:B14D6.280

A:Map position: 6

Query Match

1.3%; Score 7; DB 2; Length 181;

Best local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 EKEKKK 213

Db 108 EKEKKK 114

Search completed: July 15, 2003, 09:49:19
Job time : 45 secs



Accession	Species	Length (bp)	GC Content (%)
P39520	saccharomyces cerevisiae	1085	1.3
P78332	homo sapiens	1123	1.3

[illegible]

```

RESULT 2
TYO3_MOUSE
ID TYO3_MOUSE STANDARD: PRT: 880 AA
AC P55146
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase receptor TYRO3 precursor (P1.2.7.1.112)
DE (Tyrosine-protein kinase RSE) (Tyrosine-protein kinase DTK) (TK19-2).
CN TYRO3 OR ETK OR RSE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9419374; PubMed=7511603;
RA Mark M.R., Scadden D.T., Wang Z., Gu Q., Goddard A., Godowski P.J.;
RT "RSE, a novel receptor-type tyrosine kinase with homology to Axl/Ufo,
RT is expressed at high levels in the brain.";
RL J Biol. Chem. 269:10720-10728(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95161079; PubMed=7857657;
RA Crosier P.S., Lewis P.M., Hall L.R., Vitas M.R., Morris C.M.,
RA Beiler D.R., Wood C.R., Crosier K.R.;
RT "Isolation of a receptor tyrosine kinase (DTK) from embryonic stem
RT cells: structure, genetic mapping and analysis of expression.";
RL Growth Factors 11:125-136(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=94336210; PubMed=8058320;
RA Lai C., Gore M., Lemke G.;
RT "Structure, expression, and activity of Tyro 3, a neural adhesion-
RT related receptor tyrosine kinase.";
RL Oncogene 9:2567-2578(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=95240399; PubMed=7723626;
RA Schulz N., Faulstich C., Lee L., Zhou R.;
RT "Isolation and expression analysis of Tyro3, a murine growth factor
RT receptor tyrosine kinase preferentially expressed in adult brain";
RL Brain Res Mol. Brain Res. 28:273-280(1995).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Sasaki M.;
RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RT IN THE CENTRAL NERVOUS SYSTEM.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: ABUNDANT IN THE BRAIN AND LOWER LEVELS IN
CC OTHER TISSUES.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC AXI/UFO SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: U05683; AAA19237.1; -
DR EMBL: U18933; AAC52148.1; -
DR EMBL: X78103; CAA54995.1; ALT_INT.
DR EMBL: U18342; AAB26942.1; ALT_INT.
DR EMBL: AB000828; BAA19193.1; -
DR HSSP: P11362; IFGK.
DR MGD: MGI:104294; TYRO3.
DR InterPro: IPR000719; Euk.pkinase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003086; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR001600; Ig_Like.
DR InterPro: IPR001245; IYI_Like.
DR Pfam: PF00047; Ig_2.
DR Pfam: PF00060; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00410; Ig_Like; 1.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00219; TYRO3; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Receptor: Glycoprotein, Tyrosine-protein kinase, ATP-binding;
KW Transferase; Phosphorylation; Transmembrane; Signal; Repeat;
KW Immunoglobulin domain
FT SIGNAL 1 30
FT CHAIN 1 880
FT DOMAIN 31 419
FT TRANSMM 420 440
FT DOMAIN 441 880
FT DOMAIN 47 114
FT DOMAIN 143 209
FT DOMAIN 214 301
FT DOMAIN 305 401
FT DOMAIN 508 785
FT NP_BIND 514 522
FT BINDING 540 540
FT ACT_SITE 645 645
FT DISULF 54 107
FT DISULF 150 193
FT CARBOHYD 53 53
FT CARBOHYD 75 75
FT CARBOHYD 181 181
FT CARBOHYD 220 220
FT CARBOHYD 230 230
FT CARBOHYD 283 283
FT CARBOHYD 356 356
FT CARBOHYD 370 370
FT MOD_RES 676 676
FT CONFLICT 630 630
FT CONFLICT 811 811
SQ SEQUENCE 880 AA; 3622 MW; 3B7AC56CB01B347 C6064;
Query Match 1.58; Score 8; DB 1; Length 880;
Best Local Similarity 100.0%; Prod. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 52 DGLALHS 59
DB 240 DGLALHS 247

```

```

RESULT 3
TYO3_MOUSE
ID TYO3_MOUSE STANDARD: PRT: 880 AA
AC P55146
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase receptor TYRO3 precursor (P1.2.7.1.112)
DE (Tyrosine-protein kinase SKY).

```

GN TYROS OP SKY
 OS Rattus norvegicus (Rat)
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96104999; PubMed=7490270;
 RA Ohashi K., Honda S., Ichinomiya N., Nakamura T., Mizuno K.,
 RT "Molecular cloning and in situ localization in the brain of rat sky
 receptor tyrosine kinase";
 RI J. Biochem. 117:1267-1275(1995).
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL ADHESION PROCESSES, PARTICULARLY
 CC IN THE CENTRAL NERVOUS SYSTEM.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ABUNDANT IN THE BRAIN AND LOWER LEVELS IN
 CC OTHER TISSUES.
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES
 CC AXI/OPS SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS
 CC -----
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. The data in this file is
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR FMR1: D37880; BA07119.1: -
 DR HSSP: P11362; IEK.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR001245; Tyf_Pkinase.
 DR Pfam: PF00631; In3, 2.
 DR Pfam: PF00067; Ig_2.
 DR Pfam: PF00069; Pkinase; 1.
 DR ProDom: PDM00001; Euk_Pkinase; 1
 DR SMART: SM00063; FN3_2.
 DR SMART: SM00410; Ig_Like; 1
 DR SMART: SM00408; IGC2; 1.
 DR SMART: SM00219; TyfKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1
 KM Receptor, glycoprotein, Tyrosine-protein kinase, ATP-binding;
 KM Transferrase, Phosphorylation; Transmembrane; Signal; Repeat;
 KM Immunoglobulin domain.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 1 886 TYROSINE-PROTEIN KINASE RECEPTOR TYR-
 FT DOMAIN 31 419 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 420 440 POTENTIAL.
 FT DOMAIN 441 880 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 47 114 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 143 200 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 214 301 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 305 401 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 508 785 PROTEIN KINASE.
 FT NP_BIND 514 522 ATP (BY SIMILARITY).
 FT BINDING 540 540 ATP (BY SIMILARITY).
 FT BINDING 540 540 ATP (BY SIMILARITY).
 FT ACT_SITE 645 645 BY SIMILARITY.
 FT DISULFID 107 107 BY SIMILARITY.
 FT DISULFID 150 193 BY SIMILARITY.
 FT CARBOHYD 53 53 N-LINKED (GLCNAC).
 FT CARBOHYD 75 75 N-LINKED (GLCNAC).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC). (POTENTIAL).

PT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 676 676 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SO SEQUENCE 880 AA; 95918 MW; C3751E86A85FA5B CRC64;
 Query Match 1.5%; Score 8; DB 1; Length 880;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 52 DGLALHS 59
 DB 240 DGLALHS 247
 ID KL9_CAEEL STANDARD; PRT; 189 AA.
 AC Q95Y90;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L9
 GN L9L-9 OP P13A5.8.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Flatulida; Flatulidae;
 CC Rhabditidae; Rhabditinae; Caenorhabditis.
 CC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol NZ;
 RA Marti C., Vaudin M.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Waterson R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE L9P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AC005679; AK84469.1: -
 CC GenBank: U13A5.8; CE01340.
 DR InterPro: IPR000702; Ribosomal_L6.
 DR InterPro: IPR000554; Ribosomal_L6_2.
 DR Pfam: PF00447; Ribosomal_L6_1
 DR PROSITE: PS00700; RIBOSOMAL_L6_2; 1.
 KM Ribosomal protein.
 SV SEQUENCE 1st AA: L9LNR MW: RAFC0EAB68074F8 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 189;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 382 TAVKEKD 388
 DB 163 TAVKEKD 169
 ID YJFH_ECOLI STANDARD; PRT; 243 AA.
 AC P34290;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)


```

DR Pfam: PF01311: Bac.expert_1: 1.
DR PRINTS: PR00953: TYPE3IMPRROT.
KW Complete proteome; Protein transport; Plasmid; Virulence; Transmembrane;
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
SQ SEQUENCE 261 AA: 28451 MW: 1EE3BE9E07AD1F3 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GVLRNG 36
DB 37 GVLRNG 43

RESULT 7
ADD_RHITO STANDARD: PRT: 324 AA.
AC 098GV2:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Adenosine deaminase (EC 3.5.4.4) (Adenosine aminohydrolase).
GN ADD OR ML3163.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082930; PubMed=11214968.
RA Kaneko T., Nakamura Y., Sato S., Asami Y., Kato T., Sasamoto S.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimizu S., Sugimoto M.,
RA Takemuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti".
RL DNA Res. 7:331-338(2000).
CC -1- CATALYTIC ACTIVITY: Adenosine + H(2)O = inosine + NH(3).
CC -1- SIMILARITY: RELINGS TO THE ADENOSINE AND AMP DEAMINASE FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AP003001; BAB50114.1; -
DR InterPro: IPR001365; A/AMP_deaminase.
DR Pfam: PF00962; A_deaminase.1.
KW Hydroxylase; Nucleotide metabolism; Complete proteome.
FT ACT_SITE 186 186 POTENTIAL.
FT ACT_SITE 234 234 POTENTIAL.
FT ACT_SITE 267 267 POTENTIAL.
FT ACT_SITE 268 268 POTENTIAL.
SQ SEQUENCE 324 AA: 35247 MW: 241959E4F90EA CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 324;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 265 FVDRKTK 291
DB 306 FVDRKTK 312

RESULT 8
TRNA_NEIMB STANDARD: PRT: 362 AA.
ID TRNA_NEIMB
AC 09JYA0:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5'-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (RUMT)).
GN TRNA OR NMB1679.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN=M58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden T.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickley F.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty R.A.,
RA Mason T., Cleek A., Parksey D.S., Blair E., Cline H., Clark E.B.,
RA Cotton M.D., Hitterberg T.R., Kouri H., Qin H., Vamathevan J.,
RA Gill J., Scharf V., Maignani V., Pizarro M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon F.P., Rappaport R., Venter I.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT M58".
RL Science 287:1807-1815(2000).
CC -1- FUNCTION: catalyzes the formation of 5-methyl-uridine at position
CC 54 (M-5-U54) in all tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -1- SIMILARITY: RELINGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. TRNA
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF002518; AAF42027.1; -
DR TIGR: NMB1679; -
DR InterPro: IPR000051; SAM bind.
DR ProSITE: PS01240; TRNA_L; 1.
DR ProSITE: PS01231; TRNA_L; 2; FALSE_NEG.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 208 214 S-ADENOSYLMETHIONINE BINDING (BY
FT SIMILARITY).
FT ACT_SITE 318 318 BY SIMILARITY.
SQ SEQUENCE 362 AA: 41309 MW: 4CF18AF93E72F8EB CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 362;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 NIEANRI 88
DB 243 NIEANRI 249

RESULT 9
YGB1_METJA STANDARD: PRT: 380 AA.
AC 059075:

```

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein M1681.
 OS M1681.
 ON Methanococcus jannaschii.
 OC Archaea, Euryarchaeota, Methanococci, Methanococcales;
 OC Methanocaldococcaceae, Methanocaldococcus.
 NX NCBI_TaxID=2190.
 [1]
 RP SEQUENCE FROM N.A. 2661 / ATCC 43067;
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=9633799; PubMed=868087;
 RA Buit C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.P.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Pelech S.L.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Mortimer J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Uitterlinden T.N., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Horodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.*;
 RL Science 273:1058-1073(1996).
 CC
 CC -1- SIMILARITY: TO M.THERMOPHOTOPHILICUM MTH1681.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL institution -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 067608; AAB99702.1;
 DR TIGR: M1681;
 DR InterPro: IPR001450; 4PnAS_Ferredoxin
 DR InterPro: IPR002708; DUF39;
 DR Pfam: PF00037; Fer4_2;
 DR Pfam: PF01837; DUF39; 1;
 DR ProDom: PD01569; DUF39; 1;
 KM Hypothetical protein; Complete proteome;
 SO SEQUENCE 380 AA; 42580 MW; 03b221797974746a CRC64;
 Query Match 1.38; Score 7; FR 1; Length 380;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 336 LKRLDKD 342
 DB 253 LKRLDKD 259
 RESULT 10
 CXX1_MOUSE STANDARD; PPT: 381 AA.
 AC P33242;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Gap junction alpha-1 protein (Connexin 43) (Cx43) (Gap junction 43 kDa
 DE heart protein).
 DR GJA1 OR CXN-43.
 GN GJA1 OR CXN-43.
 OS Mus musculus (Mouse).
 OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
 OC Mammalia, Eutheria, Rodentia, Scuriognathi, Muridae, Murinae, Mus.
 NX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91217014; PubMed=1708769;
 RA Beyer E.C., Steinberg P.H.;
 RT *Evidence that the gap junction protein connexin-43 is the
 RT ATP-induced pore of mouse macrophages.*;

RL J. Biol. Chem. 266:7971-7974(1991).
 RP [2]
 RP SEQUENCE FROM N.A.
 RA Padayaratnam C.P., Morgan J.L., Lo C.W.;
 RL Submitted (SEP-1991) to the FMBL Genbank/EMBL databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVARY;
 RX MEDLINE=9126228; PubMed=906997;
 RT Nishi M., Kumar N.M., Gilula N.B.;
 RT *Developmental regulation of gap junction gene expression during
 RT mouse embryonic development.*;
 RL Dev. Biol. 146:117-130(1991).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=9229686; PubMed=1318884;
 RA Hennemann J., Sockyna T., Liechtenborg-Frater H., Jungbluth S.,
 RA Dahl E., Schwarz J., Nicholson B.J., Willock K.;
 RT *Molecular cloning and functional expression of mouse connexin40, a
 RT second gap junction gene preferentially expressed in lung.*;
 RL J. Cell Biol. 117:1299-1301(1992).
 CC
 CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
 CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXINS, THROUGH WHICH
 CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
 CC
 CC -1- FUNCTION: CONNEXIN 43 IS POSSIBLY THE ATP-INDUCED PORE OF
 CC MOUSE MACROPHAGES.
 CC
 CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
 CC
 CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. ALPHA-TYPE (GROUP 1)
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL institution -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M61896; AAA37444.1;
 DR EMBL: X61576; CAA43778.1;
 DR EMBL: M61861; AAA3027.1;
 DR EMBL: X62836; CAA44640.1;
 DR EIR: A39802; A39802;
 DR PIR: S18110; S18110;
 DR PIR: S24110; S24110;
 DR MGI: 95713; Gja1.
 DR InterPro: IPR000560; Connexin.
 DR InterPro: IPR002261; Connexin43.
 DR Pfam: PF00029; Connexin; 1;
 DR Pfam: PF03508; Connexin43; 1;
 DR PRINTS: PR00206; CONNEXIN.
 DR SMART: SM00037; CNX; 1;
 DR PROSITE: PS00407; CONNEXINS_1; 1;
 DR PROSITE: PS00408; CONNEXINS_2; 1;
 KM Gap junction; Transmembrane.
 FT INIT-MET
 FT DOMAIN 1 22
 FT TRANSSEM 23 43
 FT DOMAIN 44 75
 FT TRANSSEM 76 96
 FT DOMAIN 97 154
 FT TRANSSEM 155 175
 FT DOMAIN 176 206
 FT TRANSSEM 207 227
 FT DOMAIN 228 381
 FT TRANSSEM 319 319
 FT CONFLICT 319 319 M -> T (IN REF. 3).
 SO SEQUENCE 381 AA; 42673 MW; 96958A577979C3A5b CRC64;
 Query Match 1.38; Score 7; DR 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 VNMVEMHL 509
 11111111
 DB 120 VNMVEMHL 126

RESULT 11

ID CXAI_RAT STANDARD: PRT: 381 AA.

AC P08050;
 DT 01-AUG-1988 (rel. 08, last sequence update)
 DT 01-NOV-1988 (rel. 09, last annotation update)
 DT 16-OCT-2001 (rel. 40, last annotation update)
 DE Gap junction alpha-1 protein (connexin 43) (Cx43) (gap junction 43 kDa heart protein).
 GN GJA1 OR CXN-43.
 OS Rattus norvegicus (Rat)
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88047496; PubMed=2826492;
 RA Beyer E.C., Paul P.L., Goodenough P.A.;
 RT "Connexin43: a protein from rat heart homologous to a gap junction protein from liver."
 RL J. Cell Biol. 105:2621-2629(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9124281; PubMed=1852114;
 RA Lang L.M., Beyer E.C., Schwartz A.L., Gitting J.D.;
 RT "Molecular cloning of a rat uterine gap junction protein and analysis of gene expression during gestation."
 RL Am. J. Physiol. 260:E787-E793(1991).
 RN [3]
 RP SEQUENCE OF 1-32.
 RC TISSUE=Heart;
 RX MEDLINE=85307050; PubMed=2987225;
 RA Nicholson R.J., Gros D.R., Kent S.R.H., Hood L.E., Revel J.-P.;
 RT "The Mr 28,000 gap junction proteins from rat heart and liver are different but related."
 RL J. Biol. Chem. 260:6514-6517(1985).
 RN [4]
 RP PARTIAL SEQUENCE OF 1-15.
 RC TISSUE=Brain;
 RX MEDLINE=91348048; PubMed=1652440;
 RA Dupont E., el Aounari A., Fromaget C., Briand J.-C., Gros D.;
 RT "Affinity purification of a rat-brain junctional protein, connexin 43."
 RL Eur. J. Biochem. 200:263-270(1991).
 RN [5]
 RP DISULFIDE BONDS.
 RX MEDLINE=91337077; PubMed=1651718;
 RA Tohn S.A., Revel J.-P.;
 RT "Connexon integrity is maintained by non-covalent bonds: intramolecular disulfide bonds link the extracellular domains in rat connexin-43."
 RL Biochem. Biophys. Res. Commun. 178:1312-1318(1991).
 RN [6]
 RP TOPOLOGY
 RX MEDLINE=92167270; PubMed=1371548;
 RA Yeager M., Gilula N.B.;
 RT "Membrane topology and quaternary structure of cardiac gap junction ion channels."
 RL J. Mol. Biol. 223:929-948(1992)
 CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED PAIRS OF TRANSMEMBRANE CHANNELS. THE CONNEXONS, THROUGH WHICH MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
 CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- PTM: THERE IS AT LEAST ONE INTRAMOLECULAR DISULFIDE BOND
 CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. ALPHA-TYPE (GROUP II)

CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/submit/>
 CC or send an email to license@sib.ch).
 CC -----
 CC EMBL: X06556; CAA29855.1; -.
 CC PIP: S00532; S00532
 CC PIP: A24047; A24047
 CC InterPro: IPR002261; Connexin.
 CC InterPro: IPR002261; Connexin43.
 CC Pfam: PF00029; connexin; 1.
 CC Pfam: PF03508; Connexin43; 1.
 CC PRINTS: P000206; CONNEXIN
 CC SMART: SM00037; CNX; 1.
 CC PROSITE: PS00407; CONNEXINS_1; 1.
 CC PROSITE: PS00408; CONNEXINS_2; 1.
 CC Gap junction; Transmembrane.
 CC INIT MET 0 0
 CC DEMAIN 1 22 CYTOPLASMIC (PROBABLE).
 CC TRANSMEM 23 43 EXTRACELLULAR (PROBABLE).
 CC DOMAIN 44 75 PROBABLE.
 CC TRANSMEM 76 96 CYTOPLASMIC (PROBABLE).
 CC DOMAIN 97 154 PROBABLE.
 CC TRANSMEM 155 175 EXTRACELLULAR (PROBABLE).
 CC DEMAIN 176 206 PROBABLE.
 CC TRANSMEM 207 227 PROBABLE.
 CC DEMAIN 228 381 CYTOPLASMIC (PROBABLE).
 CC CONFLICT 1 1 G -> A (IN REF. 3).
 CC CONFLICT 15 15 A -> T (IN REF. 2).
 CC CONFLICT 27 27 V -> I (IN REF. 3).
 CC SEQUENCE 381 AA; 42900 MW; 988E907F4F99FC88 CYS64;
 CC -----

Query Match 1.38; Score 7; EB 1; Length 381;
 Best Local Similarity 100.0%; Tied. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 VNMVEMHL 509
 11111111
 DB 120 VNMVEMHL 126

RESULT 12

LN H9_HUMAN STANDARD: PRT: 388 AA.

AC U9N059; U9N070; U9BY06;
 DT 16-OCT-2001 (rel. 40, last sequence update)
 DT 16-OCT-2001 (rel. 40, last sequence update)
 DE LIM/homeobox protein Lhx9.
 GN LHX9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21291005; PubMed=11397841.
 RA Ottolenghi C., Moreira-Pilho C., Mendonca R.R., Barbieri M.,
 RA Pelous M., Berkovitz G.D., McElreavey R.;
 RT "Absence of mutations involving the LIM homeobox domain gene LHX9 in 46,XY gonadal agenesis and dysgenesis."
 PL J. Clin. Endocrinol. Metab. 86:2465-2469(2001).
 CC -1- FUNCTION: INVOLVED IN GONADAL DEVELOPMENT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAIN. THE LIM DOMAIN PRINTS 2 LING IONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See: <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: AJ277915; CAB97493.1; -;
 DR EMBL: AJ277916; CAB98128.1; ALT_SEQ.
 DR EMBL: AJ277917; CAB98128.1; JOINED.
 DR EMBL: AJ277918; CAB98128.1; JOINED.
 DR EMBL: AJ277919; CAB98128.1; JOINED.
 DR EMBL: AJ277920; CAB98128.1; JOINED.
 DR EMBL: AJ296373; CAB93174.1; -;
 DR HSP: P06601; 1FJL.
 DR Gene: HGN:14222; LHX9.
 DR MIM: 606066; -;
 DR InterPro: IPR001356; Homeobox
 DR InterPro: IPR001781; LIM.
 DR Pfam: PF00046; homeobox; 1.
 DR Pfam: PF00412; LIM; 2.
 DR ProDom: PD000010; Homeobox; 1
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00142; LIM; 2.
 DR PROSITE: PS00478; LIM_DOMAIN_1; 2.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 2.
 DR PROSITE: PS00027; HOMEBOX_1; 1
 DR PROSITE: PS50071; HOMEBOX_2; 1
 DR Homeobox: DNA binding; Nuclear protein; Repeat; LIM domain;
 DR Metal-binding; Zinc.
 DR FT DOMAIN 62 114 LIM 1
 FT DOMAIN 124 177 LIM 2.
 FT DNA_BIND 258 317 HOMEBOX.
 SO SEQUENCE 388 AA, 42986 MM, AADCBH314U7CJH06 CR6C4.

OY 387 KDGSTYC 393
 DB 105 KDGSTYC 111

Query Match 1 78, Score 7, FR 1, Length 388;
 Best Local Similarity 100 08; Pred. No. 37;
 Matches 7; conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 387 KDGSTYC 393
 DB 105 KDGSTYC 111

RESULT 13
 LHX9_MOUSE STANDARD: PRT; 388 AA.
 AC UNPUBL. U03014, U03015, U03016, U03017, U03018,
 DT 30-MAY-2000 (Ref. 49, first update)
 DT 30-MAY-2000 (Ref. 49, last sequence update)
 DT 15-JUN-2002 (Ref. 41, last annotation update)
 DE LIM/homeobox protein Lhx9.
 GN LHX9
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid:10090;
 RN [1]
 RP SEQUENCE OF 1-300 FROM N.A.
 RC STRAIN-C57BL/6;
 RX MEDLINE-9904944; PubMed-980594;
 RA Kretaux S., Rogard M., Bach I., Falli V., Besson M.-J.;
 RT "Lhx9: a novel LIM-homeobox domain gene expressed in the developing
 RT forebrain.";
 RL J. Neurosci. 19:783-793(1999).
 RT [2]
 RP SEQUENCE OF 11-388 FROM N.A.
 RC STRAIN-NIH Swiss;
 RX MEDLINE-99264291; PubMed-10330499;
 RA Hotteluzzi S., Porter F.D., Pitts A., Kumar M., Aguilinck A., Massif C.,

RA Westphal H.;
 RT "Characterization of Lhx9, a novel LIM/homeobox gene expressed by the
 RT pioneer neurons in the mouse cerebral cortex.";
 RL Mech Dev 81:193-198(1999).
 RN [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING;
 RC STRAIN-C57BL/6; Tissue-Brain;
 RX MEDLINE-2021375; PubMed-10756098;
 RA Falli V., Rogard M., Mallet M.-G., Verrier P., Kretaux S.;
 RT "Lhx9 and Lhx9alpha LIM-homeobox domain factors: genomic structure,
 RT expression patterns, chromosomal localization, and phylogenetic
 RT analysis.";
 RL Cytogenetics 64:307-317(2000).
 CC -1- FUNCTION: INVOLVED IN GONADAL DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ALPHA AND BETA (SHOWN HERE); ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BURNS 2 ZINC
 CC IONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See: <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: AF134761; AAD30110.1; -;
 DR EMBL: AF134761; AAD22009.1; -;
 DR EMBL: AJ243851; CAB59907.1; -;
 DR EMBL: AJ243852; CAB59908.1; -;
 DR EMBL: AJ243853; CAB59909.1; JOINED
 DR EMBL: AJ243854; CAB59908.1; JOINED.
 DR EMBL: AJ243855; CAB59908.1; JOINED.
 DR EMBL: AJ243856; CAB59909.1; JOINED.
 DR EMBL: AJ243857; CAB59909.1; JOINED.
 DR EMBL: AJ243858; CAB59909.1; JOINED.
 DR EMBL: AJ243859; CAB59909.1; JOINED.
 DR EMBL: AJ243860; CAB59909.1; JOINED.
 DR EMBL: AJ243861; CAB59909.1; JOINED.
 DR EMBL: AJ243862; CAB59909.1; JOINED.
 DR EMBL: AJ243863; CAB59909.1; JOINED.
 DR EMBL: AJ243864; CAB59909.1; JOINED.
 DR EMBL: AJ243865; CAB59909.1; JOINED.
 DR EMBL: AJ243866; CAB59909.1; JOINED.
 DR EMBL: AJ243867; CAB59909.1; JOINED.
 DR EMBL: AJ243868; CAB59909.1; JOINED.
 DR EMBL: AJ243869; CAB59909.1; JOINED.
 DR EMBL: AJ243870; CAB59909.1; JOINED.
 DR EMBL: AJ243871; CAB59909.1; JOINED.
 DR EMBL: AJ243872; CAB59909.1; JOINED.
 DR EMBL: AJ243873; CAB59909.1; JOINED.
 DR EMBL: AJ243874; CAB59909.1; JOINED.
 DR EMBL: AJ243875; CAB59909.1; JOINED.
 DR EMBL: AJ243876; CAB59909.1; JOINED.
 DR EMBL: AJ243877; CAB59909.1; JOINED.
 DR EMBL: AJ243878; CAB59909.1; JOINED.
 DR EMBL: AJ243879; CAB59909.1; JOINED.
 DR EMBL: AJ243880; CAB59909.1; JOINED.
 DR EMBL: AJ243881; CAB59909.1; JOINED.
 DR EMBL: AJ243882; CAB59909.1; JOINED.
 DR EMBL: AJ243883; CAB59909.1; JOINED.
 DR EMBL: AJ243884; CAB59909.1; JOINED.
 DR EMBL: AJ243885; CAB59909.1; JOINED.
 DR EMBL: AJ243886; CAB59909.1; JOINED.
 DR EMBL: AJ243887; CAB59909.1; JOINED.
 DR EMBL: AJ243888; CAB59909.1; JOINED.
 DR EMBL: AJ243889; CAB59909.1; JOINED.
 DR EMBL: AJ243890; CAB59909.1; JOINED.
 DR EMBL: AJ243891; CAB59909.1; JOINED.
 DR EMBL: AJ243892; CAB59909.1; JOINED.
 DR EMBL: AJ243893; CAB59909.1; JOINED.
 DR EMBL: AJ243894; CAB59909.1; JOINED.
 DR EMBL: AJ243895; CAB59909.1; JOINED.
 DR EMBL: AJ243896; CAB59909.1; JOINED.
 DR EMBL: AJ243897; CAB59909.1; JOINED.
 DR EMBL: AJ243898; CAB59909.1; JOINED.
 DR EMBL: AJ243899; CAB59909.1; JOINED.
 DR EMBL: AJ243900; CAB59909.1; JOINED.
 DR EMBL: AJ243901; CAB59909.1; JOINED.
 DR EMBL: AJ243902; CAB59909.1; JOINED.
 DR EMBL: AJ243903; CAB59909.1; JOINED.
 DR EMBL: AJ243904; CAB59909.1; JOINED.
 DR EMBL: AJ243905; CAB59909.1; JOINED.
 DR EMBL: AJ243906; CAB59909.1; JOINED.
 DR EMBL: AJ243907; CAB59909.1; JOINED.
 DR EMBL: AJ243908; CAB59909.1; JOINED.
 DR EMBL: AJ243909; CAB59909.1; JOINED.
 DR EMBL: AJ243910; CAB59909.1; JOINED.
 DR EMBL: AJ243911; CAB59909.1; JOINED.
 DR EMBL: AJ243912; CAB59909.1; JOINED.
 DR EMBL: AJ243913; CAB59909.1; JOINED.
 DR EMBL: AJ243914; CAB59909.1; JOINED.
 DR EMBL: AJ243915; CAB59909.1; JOINED.
 DR EMBL: AJ243916; CAB59909.1; JOINED.
 DR EMBL: AJ243917; CAB59909.1; JOINED.
 DR EMBL: AJ243918; CAB59909.1; JOINED.
 DR EMBL: AJ243919; CAB59909.1; JOINED.
 DR EMBL: AJ243920; CAB59909.1; JOINED.
 DR EMBL: AJ243921; CAB59909.1; JOINED.
 DR EMBL: AJ243922; CAB59909.1; JOINED.
 DR EMBL: AJ243923; CAB59909.1; JOINED.
 DR EMBL: AJ243924; CAB59909.1; JOINED.
 DR EMBL: AJ243925; CAB59909.1; JOINED.
 DR EMBL: AJ243926; CAB59909.1; JOINED.
 DR EMBL: AJ243927; CAB59909.1; JOINED.
 DR EMBL: AJ243928; CAB59909.1; JOINED.
 DR EMBL: AJ243929; CAB59909.1; JOINED.
 DR EMBL: AJ243930; CAB59909.1; JOINED.
 DR EMBL: AJ243931; CAB59909.1; JOINED.
 DR EMBL: AJ243932; CAB59909.1; JOINED.
 DR EMBL: AJ243933; CAB59909.1; JOINED.
 DR EMBL: AJ243934; CAB59909.1; JOINED.
 DR EMBL: AJ243935; CAB59909.1; JOINED.
 DR EMBL: AJ243936; CAB59909.1; JOINED.
 DR EMBL: AJ243937; CAB59909.1; JOINED.
 DR EMBL: AJ243938; CAB59909.1; JOINED.
 DR EMBL: AJ243939; CAB59909.1; JOINED.
 DR EMBL: AJ243940; CAB59909.1; JOINED.
 DR EMBL: AJ243941; CAB59909.1; JOINED.
 DR EMBL: AJ243942; CAB59909.1; JOINED.
 DR EMBL: AJ243943; CAB59909.1; JOINED.
 DR EMBL: AJ243944; CAB59909.1; JOINED.
 DR EMBL: AJ243945; CAB59909.1; JOINED.
 DR EMBL: AJ243946; CAB59909.1; JOINED.
 DR EMBL: AJ243947; CAB59909.1; JOINED.
 DR EMBL: AJ243948; CAB59909.1; JOINED.
 DR EMBL: AJ243949; CAB59909.1; JOINED.
 DR EMBL: AJ243950; CAB59909.1; JOINED.
 DR EMBL: AJ243951; CAB59909.1; JOINED.
 DR EMBL: AJ243952; CAB59909.1; JOINED.
 DR EMBL: AJ243953; CAB59909.1; JOINED.
 DR EMBL: AJ243954; CAB59909.1; JOINED.
 DR EMBL: AJ243955; CAB59909.1; JOINED.
 DR EMBL: AJ243956; CAB59909.1; JOINED.
 DR EMBL: AJ243957; CAB59909.1; JOINED.
 DR EMBL: AJ243958; CAB59909.1; JOINED.
 DR EMBL: AJ243959; CAB59909.1; JOINED.
 DR EMBL: AJ243960; CAB59909.1; JOINED.
 DR EMBL: AJ243961; CAB59909.1; JOINED.
 DR EMBL: AJ243962; CAB59909.1; JOINED.
 DR EMBL: AJ243963; CAB59909.1; JOINED.
 DR EMBL: AJ243964; CAB59909.1; JOINED.
 DR EMBL: AJ243965; CAB59909.1; JOINED.
 DR EMBL: AJ243966; CAB59909.1; JOINED.
 DR EMBL: AJ243967; CAB59909.1; JOINED.
 DR EMBL: AJ243968; CAB59909.1; JOINED.
 DR EMBL: AJ243969; CAB59909.1; JOINED.
 DR EMBL: AJ243970; CAB59909.1; JOINED.
 DR EMBL: AJ243971; CAB59909.1; JOINED.
 DR EMBL: AJ243972; CAB59909.1; JOINED.
 DR EMBL: AJ243973; CAB59909.1; JOINED.
 DR EMBL: AJ243974; CAB59909.1; JOINED.
 DR EMBL: AJ243975; CAB59909.1; JOINED.
 DR EMBL: AJ243976; CAB59909.1; JOINED.
 DR EMBL: AJ243977; CAB59909.1; JOINED.
 DR EMBL: AJ243978; CAB59909.1; JOINED.
 DR EMBL: AJ243979; CAB59909.1; JOINED.
 DR EMBL: AJ243980; CAB59909.1; JOINED.
 DR EMBL: AJ243981; CAB59909.1; JOINED.
 DR EMBL: AJ243982; CAB59909.1; JOINED.
 DR EMBL: AJ243983; CAB59909.1; JOINED.
 DR EMBL: AJ243984; CAB59909.1; JOINED.
 DR EMBL: AJ243985; CAB59909.1; JOINED.
 DR EMBL: AJ243986; CAB59909.1; JOINED.
 DR EMBL: AJ243987; CAB59909.1; JOINED.
 DR EMBL: AJ243988; CAB59909.1; JOINED.
 DR EMBL: AJ243989; CAB59909.1; JOINED.
 DR EMBL: AJ243990; CAB59909.1; JOINED.
 DR EMBL: AJ243991; CAB59909.1; JOINED.
 DR EMBL: AJ243992; CAB59909.1; JOINED.
 DR EMBL: AJ243993; CAB59909.1; JOINED.
 DR EMBL: AJ243994; CAB59909.1; JOINED.
 DR EMBL: AJ243995; CAB59909.1; JOINED.
 DR EMBL: AJ243996; CAB59909.1; JOINED.
 DR EMBL: AJ243997; CAB59909.1; JOINED.
 DR EMBL: AJ243998; CAB59909.1; JOINED.
 DR EMBL: AJ243999; CAB59909.1; JOINED.
 DR EMBL: AJ244000; CAB59909.1; JOINED.
 DR EMBL: AJ244001; CAB59909.1; JOINED.
 DR EMBL: AJ244002; CAB59909.1; JOINED.
 DR EMBL: AJ244003; CAB59909.1; JOINED.
 DR EMBL: AJ244004; CAB59909.1; JOINED.
 DR EMBL: AJ244005; CAB59909.1; JOINED.
 DR EMBL: AJ244006; CAB59909.1; JOINED.
 DR EMBL: AJ244007; CAB59909.1; JOINED.
 DR EMBL: AJ244008; CAB59909.1; JOINED.
 DR EMBL: AJ244009; CAB59909.1; JOINED.
 DR EMBL: AJ244010; CAB59909.1; JOINED.
 DR EMBL: AJ244011; CAB59909.1; JOINED.
 DR EMBL: AJ244012; CAB59909.1; JOINED.
 DR EMBL: AJ244013; CAB59909.1; JOINED.
 DR EMBL: AJ244014; CAB59909.1; JOINED.
 DR EMBL: AJ244015; CAB59909.1; JOINED.
 DR EMBL: AJ244016; CAB59909.1; JOINED.
 DR EMBL: AJ244017; CAB59909.1; JOINED.
 DR EMBL: AJ244018; CAB59909.1; JOINED.
 DR EMBL: AJ244019; CAB59909.1; JOINED.
 DR EMBL: AJ244020; CAB59909.1; JOINED.
 DR EMBL: AJ244021; CAB59909.1; JOINED.
 DR EMBL: AJ244022; CAB59909.1; JOINED.
 DR EMBL: AJ244023; CAB59909.1; JOINED.
 DR EMBL: AJ244024; CAB59909.1; JOINED.
 DR EMBL: AJ244025; CAB59909.1; JOINED.
 DR EMBL: AJ244026; CAB59909.1; JOINED.
 DR EMBL: AJ244027; CAB59909.1; JOINED.
 DR EMBL: AJ244028; CAB59909.1; JOINED.
 DR EMBL: AJ244029; CAB59909.1; JOINED.
 DR EMBL: AJ244030; CAB59909.1; JOINED.
 DR EMBL: AJ244031; CAB59909.1; JOINED.
 DR EMBL: AJ244032; CAB59909.1; JOINED.
 DR EMBL: AJ244033; CAB59909.1; JOINED.
 DR EMBL: AJ244034; CAB59909.1; JOINED.
 DR EMBL: AJ244035; CAB59909.1; JOINED.
 DR EMBL: AJ244036; CAB59909.1; JOINED.
 DR EMBL: AJ244037; CAB59909.1; JOINED.
 DR EMBL: AJ244038; CAB59909.1; JOINED.
 DR EMBL: AJ244039; CAB59909.1; JOINED.
 DR EMBL: AJ244040; CAB59909.1; JOINED.
 DR EMBL: AJ244041; CAB59909.1; JOINED.
 DR EMBL: AJ244042; CAB59909.1; JOINED.
 DR EMBL: AJ244043; CAB59909.1; JOINED.
 DR EMBL: AJ244044; CAB59909.1; JOINED.
 DR EMBL: AJ244045; CAB59909.1; JOINED.
 DR EMBL: AJ244046; CAB59909.1; JOINED.
 DR EMBL: AJ244047; CAB59909.1; JOINED.
 DR EMBL: AJ244048; CAB59909.1; JOINED.
 DR EMBL: AJ244049; CAB59909.1; JOINED.
 DR EMBL: AJ244050; CAB59909.1; JOINED.
 DR EMBL: AJ244051; CAB59909.1; JOINED.
 DR EMBL: AJ244052; CAB59909.1; JOINED.
 DR EMBL: AJ244053; CAB59909.1; JOINED.
 DR EMBL: AJ244054; CAB59909.1; JOINED.
 DR EMBL: AJ244055; CAB59909.1; JOINED.
 DR EMBL: AJ244056; CAB59909.1; JOINED.
 DR EMBL: AJ244057; CAB59909.1; JOINED.
 DR EMBL: AJ244058; CAB59909.1; JOINED.
 DR EMBL: AJ244059; CAB59909.1; JOINED.
 DR EMBL: AJ244060; CAB59909.1; JOINED.
 DR EMBL: AJ244061; CAB59909.1; JOINED.
 DR EMBL: AJ244062; CAB59909.1; JOINED.
 DR EMBL: AJ244063; CAB59909.1; JOINED.
 DR EMBL: AJ244064; CAB59909.1; JOINED.
 DR EMBL: AJ244065; CAB59909.1; JOINED.
 DR EMBL: AJ244066; CAB59909.1; JOINED.
 DR EMBL: AJ244067; CAB59909.1; JOINED.
 DR EMBL: AJ244068; CAB59909.1; JOINED.
 DR EMBL: AJ244069; CAB59909.1; JOINED.
 DR EMBL: AJ244070; CAB59909.1; JOINED.
 DR EMBL: AJ244071; CAB59909.1; JOINED.
 DR EMBL: AJ244072; CAB59909.1; JOINED.
 DR EMBL: AJ244073; CAB59909.1; JOINED.
 DR EMBL: AJ244074; CAB59909.1; JOINED.
 DR EMBL: AJ244075; CAB59909.1; JOINED.
 DR EMBL: AJ244076; CAB59909.1; JOINED.
 DR EMBL: AJ244077; CAB59909.1; JOINED.
 DR EMBL: AJ244078; CAB59909.1; JOINED.
 DR EMBL: AJ244079; CAB59909.1; JOINED.
 DR EMBL: AJ244080; CAB59909.1; JOINED.
 DR EMBL: AJ244081; CAB59909.1; JOINED.
 DR EMBL: AJ244082; CAB59909.1; JOINED.
 DR EMBL: AJ244083; CAB59909.1; JOINED.
 DR EMBL: AJ244084; CAB59909.1; JOINED.
 DR EMBL: AJ244085; CAB59909.1; JOINED.
 DR EMBL: AJ244086; CAB59909.1; JOINED.
 DR EMBL: AJ244087; CAB59909.1; JOINED.
 DR EMBL: AJ244088; CAB59909.1; JOINED.
 DR EMBL: AJ244089; CAB59909.1; JOINED.
 DR EMBL: AJ244090; CAB59909.1; JOINED.
 DR EMBL: AJ244091; CAB59909.1; JOINED.
 DR EMBL: AJ244092; CAB59909.1; JOINED.
 DR EMBL: AJ244093; CAB59909.1; JOINED.
 DR EMBL: AJ244094; CAB59909.1; JOINED.
 DR EMBL: AJ244095; CAB59909.1; JOINED.
 DR EMBL: AJ244096; CAB59909.1; JOINED.
 DR EMBL: AJ244097; CAB59909.1; JOINED.
 DR EMBL: AJ244098; CAB59909.1; JOINED.
 DR EMBL: AJ244099; CAB59909.1; JOINED.
 DR EMBL: AJ244100; CAB59909.1; JOINED.
 DR EMBL: AJ244101; CAB59909.1; JOINED.
 DR EMBL: AJ244102; CAB59909.1; JOINED.
 DR EMBL: AJ244103; CAB59909.1; JOINED.
 DR EMBL: AJ244104; CAB59909.1; JOINED.
 DR EMBL: AJ244105; CAB59909.1; JOINED.
 DR EMBL: AJ244106; CAB59909.1; JOINED.
 DR EMBL: AJ244107; CAB59909.1; JOINED.
 DR EMBL: AJ244108; CAB59909.1; JOINED.
 DR EMBL: AJ244109; CAB59909.1; JOINED.
 DR EMBL: AJ244110; CAB59909.1; JOINED.
 DR EMBL: AJ244111; CAB59909.1; JOINED.
 DR EMBL: AJ244112; CAB59909.1; JOINED.
 DR EMBL: AJ244113; CAB59909.1; JOINED.
 DR EMBL: AJ244114; CAB59909.1; JOINED.
 DR EMBL: AJ244115; CAB59909.1; JOINED.
 DR EMBL: AJ244116; CAB59909.1; JOINED.
 DR EMBL: AJ244117; CAB59909.1; JOINED.
 DR EMBL: AJ244118; CAB59909.1; JOINED.
 DR EMBL: AJ244119; CAB59909.1; JOINED.
 DR EMBL: AJ244120; CAB59909.1; JOINED.
 DR EMBL: AJ244121; CAB59909.1; JOINED.
 DR EMBL: AJ244122; CAB59909.1; JOINED.
 DR EMBL: AJ244123; CAB59909.1; JOINED.
 DR EMBL: AJ244124; CAB59909.1; JOINED.
 DR EMBL: AJ244125; CAB59909.1; JOINED.
 DR EMBL: AJ244126; CAB59909.1; JOINED.
 DR EMBL: AJ244127; CAB59909.1; JOINED.
 DR EMBL: AJ244128; CAB59909.1; JOINED.
 DR EMBL: AJ244129; CAB59909.1; JOINED.
 DR EMBL: AJ244130; CAB59909.1; JOINED.
 DR EMBL: AJ244131; CAB59909.1; JOINED.
 DR EMBL: AJ244132; CAB59909.1; JOINED.
 DR EMBL: AJ244133; CAB59909.1; JOINED.
 DR EMBL: AJ244134; CAB59909.1; JOINED.
 DR EMBL: AJ244135; CAB59909.1; JOINED.
 DR EMBL: AJ244136; CAB59909.1; JOINED.
 DR EMBL: AJ244137; CAB59909.1; JOINED.
 DR EMBL: AJ244138; CAB59909.1; JOINED.
 DR EMBL: AJ244139; CAB59909.1; JOINED.
 DR EMBL: AJ244140; CAB59909.1; JOINED.
 DR EMBL: AJ244141; CAB59909.1; JOINED.
 DR EMBL: AJ244142; CAB59909.1; JOINED.
 DR EMBL: AJ244143; CAB59909.1; JOINED.
 DR EMBL: AJ244144; CAB59909.1; JOINED.
 DR EMBL: AJ244145; CAB59909.1; JOINED.
 DR EMBL: AJ244146; CAB59909.1; JOINED.
 DR EMBL: AJ244147; CAB59909.1; JOINED.
 DR EMBL: AJ244148; CAB59909.1; JOINED.
 DR EMBL: AJ244149; CAB59909.1; JOINED.
 DR EMBL: AJ244150; CAB59909.1; JOINED.
 DR EMBL: AJ244151; CAB59909.1; JOINED.
 DR EMBL: AJ244152; CAB59909.1; JOINED.
 DR EMBL: AJ244153; CAB59909.1; JOINED.
 DR EMBL: AJ244154; CAB59909.1; JOINED.
 DR EMBL: AJ244155; CAB59909.1; JOINED.
 DR EMBL: AJ244156; CAB59909.1; JOINED.
 DR EMBL: AJ244157; CAB59909.1; JOINED.
 DR EMBL: AJ244158; CAB59909.1; JOINED.
 DR EMBL: AJ244159; CAB59909.1; JOINED.
 DR EMBL: AJ244160; CAB59909.1; JOINED.
 DR EMBL: AJ244161; CAB59909.1; JOINED.
 DR EMBL: AJ244162; CAB59909.1; JOINED.
 DR EMBL: AJ244163; CAB59909.1; JOINED.
 DR EMBL: AJ244164; CAB59909.1; JOINED.
 DR EMBL: AJ244165; CAB59909.1; JOINED.
 DR EMBL: AJ244166; CAB59909.1; JOINED.
 DR EMBL: AJ244167; CAB59909.1; JOINED.
 DR EMBL: AJ244168; CAB59909.1; JOINED.
 DR EMBL: AJ244169; CAB59909.1; JOINED.
 DR EMBL: AJ244170; CAB59909.1; JOINED.
 DR EMBL: AJ244171; CAB59909.1; JOINED.
 DR EMBL: AJ244172; CAB59909.1; JOINED.
 DR EMBL: AJ244173; CAB59909.1; JOINED.
 DR EMBL: AJ244174; CAB59909.1; JOINED.
 DR EMBL: AJ244175; CAB59909.1; JOINED.
 DR EMBL: AJ244176; CAB59909.1; JOINED.
 DR EMBL: AJ244177; CAB59909.1; JOINED.
 DR EMBL: AJ244178; CAB59909.1; JOINED.
 DR EMBL: AJ244179; CAB59909.1; JOINED.
 DR EMBL: AJ244180; CAB59909.1; JOINED.
 DR EMBL: AJ244181; CAB59909.1; JOINED.
 DR EMBL: AJ244182; CAB59909.1; JOINED.
 DR EMBL: AJ244183; CAB59909.1; JOINED.
 DR EMBL: AJ244184; CAB59909.1; JOINED.
 DR EMBL: AJ244185; CAB59909.1; JOINED.
 DR EMBL: AJ244186; CAB59909.1; JOINED.
 DR EMBL: AJ244187; CAB59909.1; JOINED.
 DR EMBL: AJ244188; CAB59909.1; JOINED.
 DR EMBL: AJ244189; CAB59909.1; JOINED.
 DR EMBL: AJ244190; CAB59909.1; JOINED.
 DR EMBL: AJ244191; CAB59909.1; JOINED.
 DR EMBL: AJ244192; CAB59909.1; JOINED.
 DR EMBL: AJ244193; CAB59909.1; JOINED.
 DR EMBL: AJ244194; CAB59909.1; JOINED.
 DR EMBL: AJ244195; CAB59909.1; JOINED.
 DR EMBL: AJ244196; CAB59909.1; JOINED.
 DR EMBL: AJ244197; CAB59909.1; JOINED.
 DR EMBL: AJ244198; CAB59909.1; JOINED.
 DR EMBL: AJ244199; CAB59909.1; JOINED.
 DR E

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 387 KDGSTYC 393
|||||
Db 105 KDGSTYC 111

RESULT 14
PCY2_HUMAN
ID PCY2_HUMAN STANDARD: PRT: 389 AA
AC 099447;
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Ethanolamine-phosphate cytidyltransferase (EC 2.7.7.14)
DE (Phosphorylethanolamine transferase) (GTP:phosphoethanolamine
DE cytidyltransferase).
GN PCYT2.
OS Homo sapiens (Human)
OC Eukaryota, Metazoa; Chordata, Clariata, Vertebrata, Euteleostomi,
OC Mammalia, Insecta; Primates, Catarrhini, Hominoidea, Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9723903; PubMed=9083101;
RA Nakashima A., Hosaka K., Mikawa J.;
RT Cloning of a human cDNA for GTP:phosphoethanolamine
RT cytidyltransferase by complementation in vivo of a yeast mutant."
RL J. Biol. Chem. 272:9567-9572(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strassberg R.;
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: GTP + ethanolamine phosphate = diphosphate +
CC CDP-ethanolamine.
CC -1- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS; CDP-ETHANOLAMINE SYNTHESIS;
CC SECOND STEP.
CC -1- TISSUE SPECIFICITY: STRONGEST EXPRESSION IN LIVER, HEART, AND
CC SKELETAL MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE CYTIDYLYLTRANSFERASE FAMILY
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch)
CC -----
DR FMR1: D84307; BAA12311.1; -
DR FMR1: BC000351; AA000351.1; -
DR Gene: HGNC:8756; PCYT2.
DR MIM: 602679; -
DR InterPro: IPR004821; Cyt_tran_rel.
DR InterPro: IPR004820; Cytidylyltransf.
DR Pfam: PF01467; Cytidylyltransf. 2.
DR TIGRFAMs: TIGR00125; Cyt_tran_rel. 2.
KM Transferase, Nucleocytyldyltransferase; Phospholipid biosynthe-sis
FT DOMAIN 20 194 CATALYTIC (POTENTIAL).
SQ SEQUENCE 389 AA: 43835 MW: 13FEBBAE87FEAF7F CRC64,

Query Match 1.3%; Score 7; DR 1; Length 389;
Best local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 TLETLDK 157
|||||
Db 98 TLETLDK 104

RESULT 15
ARGL_RACSU

16 ARGL_RACSU STANDARD: PRT: 400 AA.
AC P36843;
DT 01-JUN-1994 (Rel. 29, created)
DT 01-JUN-1994 (Rel. 29, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Arginine biosynthesis, bifunctional protein argJ (includes: citramate
DE N-acetyltransferase (EC 2.3.1.35) (ornithine acetyltransferase)
DE (ornithine transacetylase) (OATASE); Amino-acid acetyltransferase
DE (EC 2.3.1.1) (N-acetylglutamate synthase) (AGS)].
GN ARGJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=9429722; PubMed=8025667;
RA O'Reilly M., Devine K.M.;
RT "Sequence and analysis of the citrulline biosynthetic operon argC-F
RT from Bacillus subtilis."
RL Microbiology 143:1023-1025(1994).
RN [2]
RP SEQUENCE OF 1-80 FROM N.A.
RC STRAIN=168;
RX MEDLINE=98015415; PubMed=9459331;
RA Medina N., Vannier P., Potho R., Autret S., Lavine A., Sero S.J.;
RT Sequencing of regions downstream of addA (58 degrees) and citG (285
RT degrees) in Bacillus subtilis."
RL Microbiology 143:3305-3308(1997).
CC -1- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + L-glutamate = L-
CC ornithine + N-acetyl-L-glutamate.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + L-glutamate = CoA + N-acetyl-L-
CC glutamate.
CC -1- PATHWAY: FIRST AND FIFTH STEPS IN ARGININE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL: Z56915; CAB81544.1; -
DR EMBL: Z79580; CAB01843.1; -
DR FMR1: Y04476; CAA70639.1; -
DR FMR1: 299109; CAB12961.1; -
DR EMBL: Z99110; CAB12977.1; -
DR FIP: S78429; S78429.
DR Sublist: B01012; argJ.
DR InterPro: IPR002813; ArgJ.
DR Pfam: PF01960; ArgJ. 1.
DR Problem: P004193; ArgJ. 1.
DR TIGRFAMs: TIGR00120; ArgJ. 1.
KM Arginine biosynthesis; Transferease; Acyltransferase;
KM Methyltransferase; Complete proteome;
SQ SEQUENCE 400 AA: 44464 MW: 14445205509571 CRC64;

Query Match 1.3%; Score 7; DR 1; Length 400;
Best local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 VGSNLVK 376
|||||
Db 306 VGSNLVK 312

Search completed: July 15, 2003, 09:41:11
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2003, 09:38:25 ; Search time 26 seconds

(without alignments)
595,248 Million cell updates/sec

Title: US-10-053-192-1

Perfect score: 526
Sequence: 1 METIFNPKKEHPPIKAEVMHDLHGMFIPONDLAETTE 526

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size: 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/laa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/laa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/laa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/laa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/laa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/1/laa/backlist1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.5	205	4	US-09-385-259-3
2	8	1.5	205	4	US-09-645-370-3
3	8	1.5	188	1	US-08-445-640-12
4	8	1.5	388	3	US-08-170-558-12
5	8	1.5	388	3	US-08-447-314-12
6	8	1.5	388	3	US-08-445-661-12
7	8	1.5	533	1	US-08-488-305A-6
8	8	1.5	533	4	US-09-385-259-2
9	8	1.5	533	4	US-09-645-370-2
10	8	1.5	874	2	US-08-456-647B-6
11	8	1.5	874	2	US-08-237-401A-6
12	8	1.5	880	1	US-08-445-640-10
13	8	1.5	880	3	US-08-170-558-10
14	8	1.5	880	3	US-08-447-314-10
15	8	1.5	880	3	US-08-445-661-10
16	8	1.5	82	2	US-09-149-476-621
17	7	1.3	82	2	US-08-117-952-756
18	7	1.3	183	4	US-09-134-001C-509A
19	7	1.3	267	4	US-09-292-858A-15
20	7	1.3	479	4	US-09-442-100-13
21	7	1.3	480	4	US-08-477-928A-4
22	7	1.3	699	4	US-09-457-040B-18
23	7	1.3	771	4	US-09-090-793-8
24	7	1.3	780	1	US-08-375-709-13
25	7	1.3	780	1	US-08-752-929-13
26	7	1.3	1085	1	US-08-431-080-28
27	7	1.3	1085	2	US-08-938-514-28

28	7	1.3	1085	4	US-09-345-294-28	Sequence 28, Appl
29	7	1.3	1142	4	US-09-106-075A-89	Sequence 89, Appl
30	7	1.3	1146	4	US-08-914-999-6	Sequence 6, Appl
31	6	1.2	15	2	US-08-937-102-20	Sequence 20, Appl
32	6	1.1	15	2	US-08-937-102-21	Sequence 21, Appl
33	6	1.1	15	2	US-08-937-102-22	Sequence 22, Appl
34	6	1.1	27	3	US-08-331-625A-18	Sequence 18, Appl
35	6	1.1	27	4	US-09-494-151-18	Sequence 18, Appl
36	6	1.2	29	2	US-08-859-201-21	Sequence 21, Appl
37	6	1.1	35	4	US-08-810-009-14	Sequence 14, Appl
38	6	1.1	41	4	US-08-469-260A-382	Sequence 382, App
39	6	1.1	52	2	US-08-466-583-7	Sequence 7, Appl
40	6	1.2	52	5	PCIT-US95-07930-7	Sequence 7, Appl
41	6	1.1	54	4	US-09-187-789-49	Sequence 49, Appl
42	6	1.1	54	4	US-09-139-600-44	Sequence 44, Appl
43	6	1.1	70	2	US-08-845-256-3	Sequence 3, Appl
44	6	1.1	70	4	US-09-208-210-3	Sequence 3, Appl
45	6	1.1	89	4	US-08-936-165A-474	Sequence 474, App

ALIGNMENTS

```

RESULT 1
US-09-385-259-3
; Sequence 3, Application US/09385259
; Patent No. 6201114
; GENERAL INFORMATION:
; APPLICANT: Aquilre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; TITLE OF INVENTION: IDENTIFICATION OF GENETICALLY STATIONARY NIGHT BLINDNESS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/385,259
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-385-259-3

Query Match      1.5%, Score 8, DB 4, Length 205;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY      99 PDCKNIF 106
DB      109 PDCKNIF 116

RESULT 2
US-09-645-370-3
; Sequence 3, Application US/09645370
; Patent No. 6428958
; GENERAL INFORMATION:
; APPLICANT: Aquilre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/645,370
; EARLIER FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/385,259
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: 60/103,219
; PRIOR FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8

```

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 3
LENGTH: 205
TYPE: PRT
ORGANISM: Canis familiaris
US-09-645-370-3

Query Match 1.5%: Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 PDCKNIF 106
Db 109 PDCKNIF 116

RESULT 3
US-08-445-640-12

Sequence 12, Application US/08445640
Patent No. 5709858
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Haron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,640
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-640-12

Query Match 1.5%: Score 8; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 DGLALLS 59
Db 210 DGLALLS 217

RESULT 4
US-08-170-558-12

Sequence 12, Application US/08170558
Patent No. 6001621
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Haron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,558
FILING DATE: 20-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-170-558-12

Query Match 1.5%: Score 8; DB 3; Length 388;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 DGLALLS 59
Db 210 DGLALLS 217

RESULT 5
US-08-447-314-12

Sequence 12, Application US/08447314
Patent No. 6087144
GENERAL INFORMATION:
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Haron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,314
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-447-314-12

Query Match 1.5%, Score 8, DB 3, Length 388;
Best Local Similarity 100.0%, Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
|||||
DB 210 DGLALLHS 217

RESULT 6
US-08-445-461-12
Sequence 12, Application US/08445461
Patent No. 6096527
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,461
FILING DATE: 22-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993

ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-461-12

Query Match 1.5%, Score 8, DB 3, Length 388;
Best Local Similarity 100.0%, Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
|||||
DB 210 DGLALLHS 217

RESULT 7
US-08-488-305A-6
Sequence 6, Application US/08488305A
Patent No. 5679772
GENERAL INFORMATION:
APPLICANT: Birk, Claes Olof; Eriksson, Ulf; Peterson, Per A.
TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
Bind Thereto, Nucleic Acid Sequence Coding
Patent No. 5679772
TITLE OF INVENTION: Therefore, And Uses Thereof
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,305A
FILING DATE: 7-JUNE-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kohli, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5280.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9400
TELEFAX: (212) 688-3884
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 533 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-488-305A-6

Query Match 1.5%, Score 8, DB 1, Length 533;
Best Local Similarity 100.0%, Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PDPCKNIF 106
|||||
DB 109 PDPCKNIF 116

RESULT 8
US-09-385-259 2
Sequence 2, Application US/09385259
Patent No. 620114
GENERAL INFORMATION:
APPLICANT: Aquilite, Gustavo D.
APPLICANT: Aquilite, Gregory M.
APPLICANT: Ray, Kunal
TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
TITLE OF INVENTION: IN DOGS
FILE REFERENCE: 19603/2481
CURRENT APPLICATION NUMBER: US/09/385,259
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 60/103,219
EARLIER FILING DATE: 1998-10-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 543
TYPE: PRT
ORGANISM: Canis familiaris
US-09-385-259 2

Query Match 1.5%; Score 8; DB 4; Length 533;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 PDPCKNIF 106
|||||
DB 109 PDPCKNIF 116

RESULT 9
US-09-645-370-2
Sequence 2, Application US/09645370
Patent No. 6428958
GENERAL INFORMATION:
APPLICANT: Aquilite, Gustavo D.
APPLICANT: Acland, Gregory M.
APPLICANT: Ray, Kunal
TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
TITLE OF INVENTION: IN DOGS
FILE REFERENCE: 19603/2481
CURRENT APPLICATION NUMBER: US/09/645,370
CURRENT FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/385,259
PRIOR FILING DATE: 1999-08-30
PRIOR APPLICATION NUMBER: 60/103,219
PRIOR FILING DATE: 1998-10-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 533
TYPE: PRT
ORGANISM: Canis familiaris
US-09-645-370-2

Query Match 1.5%; Score 8; DB 4; Length 533;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 PDPCKNIF 106
|||||
DB 109 PDPCKNIF 116

RESULT 10
US-08-456-647B-6
Sequence 6, Application US/08456647B
Patent No. 5811516
GENERAL INFORMATION:
APPLICANT: Lemke Ph.D. et al., Greg E.

TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,647B
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/237,401
FILING DATE: 02-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,456
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell Ph.D., John R.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: 0721/007002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 874 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-647B-6

Query Match 1.5%; Score 8; DB 2; Length 874;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 DGLALLHS 59
|||||
DB 234 DGLALLHS 241

RESULT 11
US-08-237-401A-6
Sequence 6, Application US/08237401A
Patent No. 5837448
GENERAL INFORMATION:
APPLICANT: Lemke Ph.D. et al., Greg E.
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/237,401A
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/884,486
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Haile Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07251/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO.: 6
SEQUENCE CHARACTERISTICS:
LENGTH: 874 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-237-401A-6

Query Match 1.5%; Score 8; DB 2; Length 874;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
11111111
DB 234 DGLALLHS 241

RESULT 12
US-08-445-640-10
Sequence 10, Application US/08445640
Patent No. 5709858
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie P.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Barton, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: palin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,640
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO.: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 880 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-445-640-10

Query Match 1.5%; Score 8; DB 1; Length 880;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
11111111
DB 240 DGLALLHS 247

RESULT 13
US-08-170-558-10
Sequence 10, Application US/08170558
Patent No. 6001621
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Barton, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: palin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,558
FILING DATE: 20-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO.: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 880 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-170-558-10

Query Match 1.5%; Score 8; DB 3; Length 880;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
11111111
DB 240 DGLALLHS 247

RESULT 14
US-08-447-314-10
Sequence 10, Application US/08447314
Patent No. 6087144
GENERAL INFORMATION:
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.

```

? APPLICANT: Baron, Will F.
? TITLE OF INVENTION: Protein Tyrosine Kinases
? NUMBER OF SEQUENCES: 35
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: palin (Genentech)
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/447,314
? FILING DATE: 22-MAY-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/170558
? FILING DATE: 20-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/157563
? FILING DATE: 23-NOV-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28,616
? REFERENCE/DOCKET NUMBER: 854C1D2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 880 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
?
? US-08-447-314-10
?
? Query Match 1.5%; Score 8; DB 3; Length 880;
? Best Local Similarity 100.0%; Pred. No. 27;
? Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 52 DGLALLHS 59
? 11111111
? Db 240 DGLALLHS 247

```

```

? RESULT 15
? US-08-445-461-10
? Sequence 10, Application US/08445461
? Patent No. 6096527
? GENERAL INFORMATION:
? APPLICANT: Godowski, Paul J.
? APPLICANT: Mark, Melanie R.
? APPLICANT: Scadden, David T.
? APPLICANT: Baker, Kevin P.
? APPLICANT: Baron, Will F.
? TITLE OF INVENTION: Protein Tyrosine Kinases
? NUMBER OF SEQUENCES: 35
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: palin (Genentech)

```

```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/445,461
? FILING DATE: 22-MAY-1995
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/170558
? FILING DATE: 20-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/157563
? FILING DATE: 23-NOV-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28,616
? REFERENCE/DOCKET NUMBER: 854C3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 880 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
?
? US-08-445-461-10
?
? Query Match 1.5%; Score 8; DB 3; Length 880;
? Best Local Similarity 100.0%; Pred. No. 27;
? Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 52 DGLALLHS 59
? 11111111
? Db 240 DGLALLHS 247

```

Search completed: July 15, 2003, 09:43:17
 Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 15, 2003, 09:41:16 : Search time: 56 seconds
(without alignments)
1113.396 Million cells updated/seen

Title: US-10-053-192-1

Perfect score: 526

Sequence: 1 METFNPNKKEHPPTKAEV MHTIDKMPIDNDICAFIR 526

Scoring table: Gapop 60 0, Gapext 60 0

Searched: 445758 seqs, 116419773 residues

Word size: 0

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Published_Applications_AA*

1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUB_PEP.*
5: /cgn2_6/ptodata/1/pubpaa/US05_PUB_PEP.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUB_PEP.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUB_PEP.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUB_PEP.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUB_PEP.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUB_PEP.*

Prod. No. is the number of results provided by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	526	100	6	US-10-053-192-1	Sequence 1, Appl 1
2	506	96.2	506	US-10-053-192-4	Sequence 4, Appl 1
3	18	3.4	18	US-10-053-192-6	Sequence 6, Appl 1
4	18	3.4	18	US-10-053-192-7	Sequence 7, Appl 1
5	8	1.5	8	US-10-053-192-3	Sequence 3, Appl 1
6	8	1.5	386	US-09-985-675-5	Sequence 5, Appl 1
7	8	1.5	388	US-09-223-428-17	Sequence 17, Appl 1
8	1.5	529	9	US-10-053-192-5	Sequence 5, Appl 1
9	1.5	580	9	US-09-928-457-37	Sequence 37, Appl 1
10	1.5	850	10	US-09-945-675-2	Sequence 2, Appl 1
11	8	1.5	874	US-09-158-722-6	Sequence 6, Appl 1
12	8	1.5	874	US-09-985-675-1	Sequence 1, Appl 1
13	8	1.5	880	US-09-223-428-10	Sequence 10, Appl 1
14	7	1.3	23	US-09-809-391-621	Sequence 621, Appl 1
15	7	1.3	32	US-09-804-761-34560	Sequence 34560, Appl 1
16	7	1.3	53	US-09-789-561-118	Sequence 118, Appl 1
17	7	1.3	53	US-09-789-561-148	Sequence 148, Appl 1
18	7	1.3	53	US-09-789-561-149	Sequence 149, Appl 1
19	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1

20	7	1.3	124	US-09-216-444-64	Sequence 64, Appl 1
21	7	1.3	102	US-09-171-161A-100	Sequence 160, Appl 1
22	7	1.3	256	US-10-106-609-4920	Sequence 4920, Appl 1
23	7	1.3	358	US-09-738-626-5295	Sequence 5295, Appl 1
24	7	1.3	410	US-09-983-931-7	Sequence 2, Appl 1
25	7	1.3	771	US-10-331-061-8	Sequence 8, Appl 1
26	7	1.3	860	US-10-128-714-7640	Sequence 7640, Appl 1
27	7	1.3	869	US-10-128-714-3239	Sequence 3239, Appl 1
28	7	1.3	869	US-09-823-376-9	Sequence 9, Appl 1
29	7	1.3	1028	US-10-153-668-192	Sequence 192, Appl 1
30	7	1.3	1129	US-10-158-714-9239	Sequence 9239, Appl 1
31	7	1.3	1146	US-09-832-293-10	Sequence 10, Appl 1
32	7	1.3	1146	US-09-994-485-6	Sequence 6, Appl 1
33	7	1.3	1146	US-10-211-088-271	Sequence 271, Appl 1
34	6	1.1	15	US-10-052-831-127	Sequence 127, Appl 1
35	6	1.1	19	US-09-553-223-49	Sequence 49, Appl 1
36	6	1.1	25	US-09-258-031B-1	Sequence 1, Appl 1
37	6	1.1	25	US-09-974-879-605	Sequence 605, Appl 1
38	6	1.1	25	US-09-305-736-606	Sequence 606, Appl 1
39	6	1.1	25	US-09-929-955-36	Sequence 36, Appl 1
40	6	1.1	25	US-09-972-484-18	Sequence 18, Appl 1
41	6	1.1	27	US-10-323-282-534	Sequence 534, Appl 1
42	6	1.1	29	US-09-910-562-7	Sequence 7, Appl 1
43	6	1.1	34	US-10-102-806-659	Sequence 659, Appl 1
44	6	1.1	34	US-09-776-490-14	Sequence 14, Appl 1
45	6	1.1	45	US-09-776-490-14	Sequence 14, Appl 1

ALIGNMENTS

RESULT 1	US-10-053-192-1	100.0%	Prod. No. 0	Length: 526
1	Sequence 1, Appl 1	100.0%	Prod. No. 0	Length: 526
2	Published ID No. 8230308730A1	100.0%	Prod. No. 0	Length: 526
3	GENERAL INFORMATION:	100.0%	Prod. No. 0	Length: 526
4	APPLICANT: BACHMANN, Heinrich	100.0%	Prod. No. 0	Length: 526
5	APPLICANT: BRIDGEMAN, Roland	100.0%	Prod. No. 0	Length: 526
6	APPLICANT: FRIEDLEIN, Arno M	100.0%	Prod. No. 0	Length: 526
7	APPLICANT: WITZ, Gabriele M	100.0%	Prod. No. 0	Length: 526
8	APPLICANT: WISS, Adrian	100.0%	Prod. No. 0	Length: 526
9	APPLICANT: WISS, Markus	100.0%	Prod. No. 0	Length: 526
10	TITLE OF INVENTION: BETA-BETA-CAROTENE 15,15'-HYDROXYENANES NUCLEIC ACID	100.0%	Prod. No. 0	Length: 526
11	TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE	100.0%	Prod. No. 0	Length: 526
12	FILE REFERENCE: E-PATENT 15,15'-HYDROXYENANES	100.0%	Prod. No. 0	Length: 526
13	CURRENT APPLICATION NUMBER: US/10/053,192	100.0%	Prod. No. 0	Length: 526
14	PRIOR APPLICATION NUMBER: 2002-01-15	100.0%	Prod. No. 0	Length: 526
15	PRIOR FILING DATE: 1999-02-22	100.0%	Prod. No. 0	Length: 526
16	NUMBER OF SEQ ID NOS: 10	100.0%	Prod. No. 0	Length: 526
17	SOFTWARE: PatentIn Ver. 2.1	100.0%	Prod. No. 0	Length: 526
18	SEQ ID NO 1	100.0%	Prod. No. 0	Length: 526
19	TYPE: PRT	100.0%	Prod. No. 0	Length: 526
20	ORGANISM: CHICKEN	100.0%	Prod. No. 0	Length: 526
21	US-10-053-192-1	100.0%	Prod. No. 0	Length: 526
22	Query Match	100.0%	Prod. No. 0	Length: 526
23	Best Local Similarity	100.0%	Prod. No. 0	Length: 526
24	Matches: 526	100.0%	Prod. No. 0	Length: 526
25	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
26	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
27	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
28	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
29	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
30	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
31	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
32	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
33	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
34	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
35	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
36	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
37	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
38	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
39	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
40	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
41	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
42	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
43	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
44	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526
45	1 METFNPNKKEHPPTKAEVGGQLPTMDGVLPLNGSGMHTIGSTYNNHFDLALLHSF 60	100.0%	Prod. No. 0	Length: 526

QY 181 ILNMGTSIVDKGRKYVLEKIPSSVPEKEKKKSCFKHLEVYCSIPSRSLLOPSYHSGI 240
DB 181 ILNMGTSIVDKGRKYVLEKIPSSVPEKEKKKSCFKHLEVYCSIPSRSLLOPSYHSGI 240
QY 241 TENYIVFIEQPFKLDIYKLTATAYIRGVNMASSCLSFHKEDKTMWFHFDRTKKEVSTKFT 300
DB 241 TENYIVFIEQPFKLDIYKLTATAYIRGVNMASSCLSFHKEDKTMWFHFDRTKKEVSTKFT 300
QY 301 DALVLYHHINAYEEDGHVFDIYAYRDNLSLYDMFYLLKLDKDFEVNKKLTSLPTCKRFVY 360
DB 301 DALVLYHHINAYEEDGHVFDIYAYRDNLSLYDMFYLLKLDKDFEVNKKLTSLPTCKRFVY 360
QY 361 PLOYDKAEVGSNLVLPSTATAVKEKDSICYCPILLCEGIELPRVNDYNGKRYKYY 420
DB 361 PLOYDKAEVGSNLVLPSTATAVKEKDSICYCPILLCEGIELPRVNDYNGKRYKYY 420
QY 421 ATEVQMSVPPTKIAKLNVQTEKYLHWGEHCHWSEPTFVSPDAREDEGVVLLCVVSE 480
DB 421 ATEVQMSVPPTKIAKLNVQTEKYLHWGEHCHWSEPTFVSPDAREDEGVVLLCVVSE 480
QY 481 PNKAPFLILIDAKTFKELGRATVNVEMHLLHGMFLPNDLGAETE 526
DB 481 PNKAPFLILIDAKTFKELGRATVNVEMHLLHGMFLPNDLGAETE 526

RESULT 2
US-10-053-192-4

Sequence 4, Application US/10053192
Publication No. US20030087336A1
GENERAL INFORMATION:
APPLICANT: BACHMANN, Heinrich
APPLICANT: BRUGGER, Roland
APPLICANT: FRIEDELIN, Arno M
APPLICANT: WIRTZ, Gabriele M
APPLICANT: WOGGON, Wolf-Dietrich
APPLICANT: WYSS, Adrian
TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...
CURRENT APPLICATION NUMBER: 04/10/053, 192
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 103182.0
PRIOR FILING DATE: 1999-02-22
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 4
LENGTH: 506
TYPE: PRT
ORGANISM: CHICKEN
US-10-053-192-4

Query Match 96.2%; Score 506; DB 9, Length 506;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 10 EEHPEPIKAEVQGLPPIWIGVLLRNGPCMHITIGTKYHMFQGLALLHSFTFKNGEYVY 69
DB 1 EEHPEPIKAEVQGLPPIWIGVLLRNGPCMHITIGTKYHMFQGLALLHSFTFKNGEYVY 69
QY 70 RSKYLRSDYNCNIEANRIIVSEFGIMAYPDPCKNIFAKAFSYLSHTIPEFTDNLINIM 129
DB 61 RSKYLRSDYNCNIEANRIIVSEFGIMAYPDPCKNIFAKAFSYLSHTIPEFTDNLINIM 120
QY 140 KTGDYVATSETNFIKIDPQLETLDKVDYSKYAVANLATSHPRYDSAGNILLNGTSIV 189
DB 121 KTGDYVATSETNFIKIDPQLETLDKVDYSKYAVANLATSHPRYDSAGNILLNGTSIV 180
QY 190 DKGRKYVLEKIPSSVPEKEKKKSCFKHLEVYCSIPSRSLLOPSYHSGIENYIVIE 249
DB 181 DKGRKYVLEKIPSSVPEKEKKKSCFKHLEVYCSIPSRSLLOPSYHSGIENYIVIE 240

QY 250 QPFKLDIYKLTATAYIRGVNMASSCLSFHKEDKTMWFHFDRTKKEVSTKRYTIALVLYHHI 409
DB 241 QPFKLDIYKLTATAYIRGVNMASSCLSFHKEDKTMWFHFDRTKKEVSTKRYTIALVLYHHI 400
QY 310 NAYEEDGHVFDIYAYRDNLSLYDMFYLLKLDKDFEVNKKLTSLPTCKRFVNDYNGKRYKYY 469
DB 301 NAYEEDGHVFDIYAYRDNLSLYDMFYLLKLDKDFEVNKKLTSLPTCKRFVNDYNGKRYKYY 460
QY 370 VGSNLVLPSTATAVKEKDSICYCPILLCEGIELPRVNDYNGKRYKYYATEVQMSVP 429
DB 361 VGSNLVLPSTATAVKEKDSICYCPILLCEGIELPRVNDYNGKRYKYYATEVQMSVP 420
QY 430 PTKIAKLNVQTEKYLHWGEHCHWSEPTFVSPDAREDEGVVLLCVVSEPNKAPFLIL 489
DB 421 PTKIAKLNVQTEKYLHWGEHCHWSEPTFVSPDAREDEGVVLLCVVSEPNKAPFLIL 480
QY 490 LDKTFKELGRATVNVEMHLLHGMFLPNDLGAETE 526
DB 481 LDKTFKELGRATVNVEMHLLHGMFLPNDLGAETE 526

RESULT 3
US-10-053-192-6

Sequence 6, Application US/10053192
Publication No. US20030087336A1
GENERAL INFORMATION:
APPLICANT: BACHMANN, Heinrich
APPLICANT: BRUGGER, Roland
APPLICANT: FRIEDELIN, Arno M
APPLICANT: WIRTZ, Gabriele M
APPLICANT: WOGGON, Wolf-Dietrich
APPLICANT: WYSS, Adrian
TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...
CURRENT APPLICATION NUMBER: 04/10/053, 192
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 103182.0
PRIOR FILING DATE: 1999-02-22
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 6
LENGTH: 18
TYPE: PRT
ORGANISM: CHICKEN
US-10-053-192-6

Query Match 3.4%; Score 18; DB 9, Length 18;
Best Local Similarity 100.0%; Pred. No. 4, 66-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NKEHPEPIKAEVQGLP 25
DB 1 NKEHPEPIKAEVQGLP 18

RESULT 4
US-10-053-192-7

Sequence 7, Application US/10053192
Publication No. US20030087336A1
GENERAL INFORMATION:
APPLICANT: BACHMANN, Heinrich
APPLICANT: BRUGGER, Roland
APPLICANT: FRIEDELIN, Arno M
APPLICANT: WIRTZ, Gabriele M
APPLICANT: WOGGON, Wolf-Dietrich
APPLICANT: WYSS, Adrian
TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...
CURRENT APPLICATION NUMBER: 04/10/053, 192

US-10-053-192-7

Query Match 3.4%; Score 18; DB 9; Length 18;
Best Local Similarity 100.0%; Pred No. 3; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 NKEHEPEIKAEVQGLP 18

RESULT 5

US-10-053-192-3

Sequence 3, Application US/10053192
Publication No. US20030087336A1
GENERAL INFORMATION:
APPLICANT: BACHMANN, Heinrich
APPLICANT: BRUGGER, Roland M
APPLICANT: FRIEDLEIN, Arno M
APPLICANT: WIRTS, Gabriele M
APPLICANT: WOSGON, Wolf-Dietrich
APPLICANT: WYSS, Adrian
APPLICANT: WYSS, Markus
TITLE OF INVENTION: BETA-BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
FILE REFERENCE: B,B-CAROTENE 15,15'-DIOXYGENASES, ...
CURRENT APPLICATION NUMBER: US/10/053,192
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 103382 0
PRIOR FILING DATE: 1999-02-22
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
LENGTH: 8
TYPE: PRT
ORGANISM: CHICKEN

US-10-053-192-3

Query Match 1.5%; Score 8; DB 9; Length 8;
Best Local Similarity 100.0%; Pred No. 4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AEVQGLP 25

Db 1 AEVQGLP 8

RESULT 6

US-09-985-675-5

Sequence 5, Application US/09985675
Patent No. US20020137168A1
GENERAL INFORMATION:
APPLICANT: CROSIER, PHILIP S.
APPLICANT: CROSIER, KATHRYN F
TITLE OF INVENTION: DEVELOPMENTAL TYROSINE KINASES AND
THEIR LIGANDS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYTE P.C.
STREET: 1100 NORTH GLENN ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U S A
ZIP: 22201-4714

US-09-985-675-5

Query Match 1.5%; Score 8; DB 10; Length 386;
Best Local Similarity 100.0%; Pred No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 210 DGLALHS 217

RESULT 7

US-09-923-490-12

Sequence 12, Application US/0923490
Patent No. US20020147325A1
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-FOS/MS-FOS
SOFTWARE: PatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/985,675
FILING DATE: 05-Nov-98
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/505,241
FILING DATE: 15-Aug-1995
ATTORNEY/AGENT INFORMATION:
NAME: MITCHELL, LEONARD C.
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 175-19
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 386 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-985-675-5

US-09-985-675-5

Query Match 1.5%; Score 8; DB 10; Length 386;
Best Local Similarity 100.0%; Pred No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 210 DGLALHS 217

RESULT 7

US-09-923-490-12

Sequence 12, Application US/0923490
Patent No. US20020147325A1
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-FOS/MS-FOS
SOFTWARE: PatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/985,675
FILING DATE: 05-Nov-98
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/505,241
FILING DATE: 15-Aug-1995
ATTORNEY/AGENT INFORMATION:
NAME: MITCHELL, LEONARD C.
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 175-19
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 386 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-985-675-5

TELEPHONE: 415/225-1896
 TELEFAX: 415/952-9881
 TELE: 910/371-7168
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 388 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-09-223-490-12

Query Match 1.5%: Score 8; DB 10; Length 388;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
 |||||
 DB 210 DGLALLHS 217

RESULT 8
 US-10-053-192-5
 : Sequence 5, Application US/10053192
 : Publication No. US20030087336A1
 : GENERAL INFORMATION:
 : APPLICANT: BACHMANN, Heinrich
 : APPLICANT: BROGGER, Roland
 : APPLICANT: FRIEDLEIN, Arno M
 : APPLICANT: WIRTZ, Gabriele M
 : APPLICANT: WOGGON, Wolf-Dietrich
 : APPLICANT: WYSS, Adrian
 : TITLE OF INVENTION: BETA-HETA-CAPOTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
 : FILE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
 : TITLE REFERENCE: R,B-CAPOTENE 15,15'-DIOXYGENASES,...
 : CURRENT APPLICATION NUMBER: US/10/053,192
 : CURRENT FILING DATE: 2002-01-15
 : PRIOR APPLICATION NUMBER: 103382.0
 : PRIOR FILING DATE: 1999-02-22
 : NUMBER OF SEQ ID NOS: 10
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 5
 : LENGTH: 529
 : TYPE: PRT
 : ORGANISM: HOVINE
 US-10-053-192-5

Query Match 1.5%: Score 8; DB 9; Length 529;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PDPCKNIF 106
 |||||
 DB 94 PDPCKNIF 101

RESULT 9
 US-09-928-457-37
 : Sequence 37, Application US/09928457
 : Patent No. US20020164603A1
 : GENERAL INFORMATION:
 : APPLICANT:
 : TITLE OF INVENTION: DNA, specific proteins and peptides
 : TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
 : TITLE OF INVENTION: for obtaining them and their biological application.
 : NUMBER OF SEQUENCES: 99
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30 (OEB)
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/928,457
 : FILING DATE: 2001-08-14

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/214,759
 FILING DATE: 199-12-10
 INFORMATION FOR SEQ ID NO: 37:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 580 acids amin,s
 TYPE: acid amin,
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Protein
 LOCATION: 1..580
 US-09-928-457-37

Query Match 1.5%: Score 8; DB 9; Length 580;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VNYDYNCK 414
 |||||
 DB 325 VNYDYNCK 332

RESULT 10
 US-09 985 675-2
 : Sequence 2, Application US/09985675
 : Patent No. US2002017168A1
 : GENERAL INFORMATION:
 : APPLICANT: CROSTIER, PHILIP S.
 : TITLE OF INVENTION: DEVELOPMENTAL TYPE-SINE KINASES AND
 : THEIR LIGANDS
 : NUMBER OF SEQUENCES: 16
 : CURRENT INVENTOR ADDRESS:
 : ADDRESSEE NIXON & VANDERHUYE P.C.
 : STREET: 1100 NORTH GLEBE ROAD
 : CITY: ARLINGTON
 : STATE: VIRGINIA
 : COUNTRY: U.S.A.
 : ZIP: 22201-4714
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/985,675
 : FILING DATE: 05-NOV-02 US2002017168A1-2001
 : CLASSIFICATION: <unknown>
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/505,241
 : FILING DATE: 16-AUG-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: MITCHARD, LEONARD C.
 : REGISTRATION NUMBER: 29,009
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 816-4000
 : TELEFAX: (703) 816-4100
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 850 AMINO ACIDS
 : TYPE: AMINO ACID
 : TOPOLOGY: LINEAR
 : MOLECULE TYPE: PROTEIN
 : SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Query Match 1.5%: Score 8; DB 10; Length 850;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59
DB 210 DGLALLHS 217

RESULT 11

US-09-158-722-6
Sequence 6, Application US/09158722
Publication No. US2003001348A1
GENERAL INFORMATION:
APPLICANT: Lemke Ph.D. et al., Greg E.
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 425 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/158,722
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/456,647
FILING DATE: 02-JUN-1995
APPLICATION NUMBER: US 08/237,401
FILING DATE: 02-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,486
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell Ph.D., John R.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: 07/251/007902
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 874 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-158-722-6

Query Match 1.58, Score 8, DB 3, Length 874,
Best Local Similarity 100.0%, Pred. No. 58;
Matches 8; Conservative 0, Mismatches 0, Indels 0, Gaps 0,

QY 52 DGLALLHS 59
DB 234 DGLALLHS 241

RESULT 12

US-09-985-675-1
Sequence 1, Application US/09985675
Patent No. US20030137168A1
GENERAL INFORMATION:
APPLICANT: CROSIER, PHILIP S.
CROSIER, KATHRYN E.
TITLE OF INVENTION: DEVELOPMENTAL TYROSINE KINASES AND
THEIR LIGANDS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHIE P.C.

STREET: 1100 NORTH GILFRE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/985,675
FILING DATE: 05-NOV-2003/168A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/555,241
FILING DATE: 16-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: MITCHELL, LEONARD C.
REGISTRATION NUMBER: 39,009
REFERENCE/DOCKET NUMBER: 175-19
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 874 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-985-675-1

Query Match 1.58, Score 8, DB 10, Length 874;
Best Local Similarity 100.0%, Pred. No. 58;
Matches 8; Conservative 0, Mismatches 0, Indels 0, Gaps 0,

QY 52 DGLALLHS 59
DB 234 DGLALLHS 241

RESULT 13

US-09-223-490-10
Sequence 10, Application US/09223490
Patent No. US20020147325A1
GENERAL INFORMATION:
APPLICANT: Godowski, Paul T.
APPLICANT: Mark, Melanie R.
APPLICANT: Spadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Barron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 Inch, 360 Kb Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/223,490
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/10/0,568
FILING DATE:

```

ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28 616
: REFERENCE/DOCKET NUMBER: 854C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1896
: TELEFAX: 415/952-9881
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 880 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-09-223-490 10

Query Match: 1.58; Score 8; DB 10; Length 880;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DIGITALHS 59
DB 240 DIGITALHS 247

RESULT 14
US-09-809-391-621
: Sequence 621, Application US/09809391
: Publication No. US2003004961A1
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: 186 Human Secreted proteins
: FILE REFERENCE: P2002P2
: CURRENT APPLICATION NUMBER: US/09/809,391
: CURRENT FILING DATE: 2001-03-16
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 761
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 621
: LENGTH: 23
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-809-391-621

Query Match: 1.38; Score 7; DB 9; Length 23;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 KIPSSVP 206
DB 12 KIPSSVP 18

RESULT 15
US-09-864-761-34560
: Sequence 34560, Application US/09864761
: Patent No. US20020284763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Acemica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6

```

```

: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/808,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 34560
: LENGTH: 32
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC054814.2
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL: 4.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL: 3.8
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
: OTHER INFORMATION: EXPRESSED IN BELA, SIGNAL: 4.5
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL: 4.8
: OTHER INFORMATION: EXPRESSED IN HRI 100, SIGNAL: 4.1
: OTHER INFORMATION: EXPRESSED IN HONE MARROW, SIGNAL = 5.1
: OTHER INFORMATION: EST HUMAN HIT: AJ120419.1, EVALUATE 2.00e-10
: OTHER INFORMATION: SWISSPROT HIT: P70501, EVALUATE 2.90e-02
US-09-864-761-34560

```

```

Query Match: 1.38; Score 7; DB 10; Length 32;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 VAVNLAT 170
DB 23 VAVNLAT 29

Search completed: July 15, 2003, 09:50:21
top time: 56 secs

```

GenCorp version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

CM protein - protein search, using sw model

Run on: July 15, 2003, 04:14:43 ; Search time: 6 seconds
(without alignments)

1275 Million cell updates/sec

Title: US-10-053-192-1

Sequence: 1 M E T T F N K E E H P P I A A V M H L L H G M F I P Q N L S A E T E 546

Scoring table: **BLOSUM62**

Searched: 671540 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671,681

Maximum pk seq length: 20000000000

Post-processing - Minimum Match of 100%

Listing first 45 summaries

Database :

- 1: sp.archae.*
- 2: sp.bacteri.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mmc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.tooth.*
- 12: sp.virus.*
- 13: sp.unclassified.*
- 14: sp.virus.*
- 15: sp.bacteriaph.*
- 16: sp.bacteriaph.*
- 17: sp.bacteriaph.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARY

Query No.	Score	Query Match	Length	FR	IF	Insertion
1	2825	100.0	546	14	Q91993	Q91993 gallus gall
2	2124	75.2	556	11	Q91056	Q91056 mus muscul
3	2121	75.1	566	11	Q96891	Q96891 mus muscul
4	2118	75.0	566	11	Q91815	Q91815 rattus norv
5	2107	74.8	547	4	Q94NV6	Q94NV6 homo sapien
6	2104	74.6	547	4	Q94NV5	Q94NV5 homo sapien
7	1215	60.7	516	13	Q90NM14	Q90NM14 brachydanio
8	1142	40.4	549	13	Q90NM14	Q90NM14 brachydanio
9	1115	40.5	545	4	Q961V8	Q961V8 rattus norv
10	1112.5	39.4	532	11	Q94YN1	Q94YN1 mus sapien
11	1102.5	39.0	539	4	Q961V5	Q961V5 mus muscul
12	1102.5	39.0	556	4	Q96NV7	Q96NV7 homo sapien
13	1077	38.1	533	11	Q91Z05	Q91Z05 mus muscul
14	1075	38.1	533	4	Q15618	Q15618 homo sapien
15	1072	37.6	533	11	Q976276	Q976276 rattus norv
16	1071	37.5	533	6	Q976274	Q976274 rattus norv

17	1070	37.9	533	6	Q9X171	Q9X171 cerropitheci
16	1070	37.9	533	6	Q9X561	Q9X561 bos laurus
19	1069	37.8	533	13	Q9X322	Q9X322 allis galli
20	1068	37.8	533	6	Q28175	Q28175 bos laurus
21	1068	37.8	533	6	Q91V68	Q91V68 canis famli
22	1064	37.7	533	13	Q9Y125	Q9Y125 abysstoma t
23	993.5	35.2	483	11	Q9VH12	Q9VH12 mus musculu
24	840.5	29.4	620	5	Q9VPS2	Q9VPS2 dirosophila
25	838.5	29.7	620	5	Q9QMK9	Q9QMK9 dirosophila
26	813.5	28.8	544	5	Q9Q2M4	Q9Q2M4 canoribadi
27	778.5	27.6	556	5	Q9QX19	Q9QX19 canoribadi
28	652.5	23.1	293	11	Q9CFC3	Q9CFC3 mus musculu
29	541	19.2	570	10	Q98V25	Q98V25 arabidopsis
30	445	15.8	552	19	Q93VU5	Q93VU5 cryta sativa
31	437	15.5	416	10	Q9M079	Q9M079 arabidopsis
32	395	14.0	460	16	Q74324	Q74324 arabidopsis
33	376.5	13.3	497	16	Q8Y1B4	Q8Y1B4 arabidopsis
34	312	11.0	442	2	Q8A866	Q8A866 arabidopsis
35	311	11.0	536	16	Q95572	Q95572 arabidopsis
36	310	11.0	538	10	Q91613	Q91613 arabidopsis
37	301.5	10.7	542	16	Q941F2	Q941F2 arabidopsis
38	280	9.9	480	16	Q74370	Q74370 arabidopsis
39	279	9.5	472	16	Q81XV3	Q81XV3 arabidopsis
40	276	9.8	605	10	Q92453	Q92453 arabidopsis
41	269	9.5	790	5	Q96904	Q96904 arabidopsis
42	269	9.5	842	5	Q950P5	Q950P5 arabidopsis
43	264.5	9.4	504	16	Q94P46	Q94P46 arabidopsis
44	264.5	9.4	472	16	Q9XVX8	Q9XVX8 arabidopsis
45	264	9.3	604	10	Q9M329	Q9M329 arabidopsis

ALIGNMENT

RESULT 1

ID	091993:	PRELIMINARY:	PRT:	526 AA.
DT	01-OCT-2000 (TREMBL)	15, Created		
DT	01-OCT-2000 (TREMBL)	15, Last sequence update		
DT	01-DEC-2001 (TREMBL)	15, Last annotation update		
DE	Beta-carotene 15,15'-dioxygenase (EC 1.11.11.21).			
GN	BCDO.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Ciconiata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae			
OC	Gallus.			
OX	NCBI_TaxID=9031:			
OX	11			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LSL LOHMANN; TISSUE=EMBRYO;			
RC	MEDLINE=20261261; PubMed=10799297;			
PA	Wess A, Wittz G M, Worgon W D, Brugger P, Wysz M, Friedlein A.			
RA	Beckmann H, Hünziker W.			
RT	Cloning and expression of beta-carotene-15,15'-dioxygenase.			
PL	Biotechnol. Biophys. Res. Commun. 271:334-346(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LSL LOHMANN; TISSUE=DUODENUM;			
RC	MEDLINE=211436; PubMed=1127856;			
PA	Wess A, Wittz G M, Worgon W D, Brugger P, Wysz M, Friedlein A.			
RA	Riss G, Beckmann H, Hünziker W.			
RT	Expression pattern and localization of beta-carotene 15,15'-			
RT	dioxygenase in different tissues.			
PL	Biochem. J. 354:521-529(2001).			
DP	EMBL: AJ271866; CAB90825.1;			
DE	InterPro: IPI004294, RPB65.			
DE	Pfam: PF03085; RPB65.			
KM	Dioxygenase: Oxidoreductase.			
SC	SEQUENCE 526 AA; 60384 MW; 8265EAP75F45H CDS4;			
ENTRY NAME	100.00; SCORE 28.45; MH 14; LENGTH 526;			
Post-local Similarity	100.00; PROD 88.4; ID 202;			

	Matches	52%, Conservative	0;	Mismatches	0	Indels	0;	Gaps	0;
QY	1	MEPIFNKNEHHPEPKAFVQGLPTMIGVYLLANGQCHMTIGDTYKNHMFQDALLHSP	60						
Db	1	MEPIFNKNEHHPEPKAFVQGLPTMIGVYLLANGQCHMTIGDTYKNHMFQDALLHSP	60						
QY	61	TFPNNGEYVRSKYI KSPITVNCNIEANPIVSEEGTMAVPPCKNIFAKAFSYLSHTIPSF	120						
Db	61	TFPNNGEYVRSKYI KSPITVNCNIEANPIVSEEGTMAVPPCKNIFAKAFSYLSHTIPSF	120						
QY	121	TDMCLINIMKTDGDIYATISENFIKKIDPOTLETLDDVYSKYVAANLATSHPHYDSAGN	180						
Db	121	TDMCLINIMKTDGDIYATISENFIKKIDPOTLETLDDVYSKYVAANLATSHPHYDSAGN	180						
QY	181	ILNMGSTIYDKGKTKVLPKIPSSVPEKKEKKSCKFHLFWCSIPSSILLQPSYHSGI	240						
Db	181	ILNMGSTIYDKGKTKVLPKIPSSVPEKKEKKSCKFHLFWCSIPSSILLQPSYHSGI	240						
QY	241	TENVIVFIEIQPKRLDVKLATAYIRGVMAASCLSFHREDEKTFWFFVDRKTKKEVSTFPT	300						
Db	241	TENVIVFIEIQPKRLDVKLATAYIRGVMAASCLSFHREDEKTFWFFVDRKTKKEVSTFPT	300						
QY	301	DALVLTNHNAAEEGCHVYFDIVARONSILDMETYLKLDKDFVNNKLTSLPTCKRPVY	360						
Db	301	DALVLTNHNAAEEGCHVYFDIVARONSILDMETYLKLDKDFVNNKLTSLPTCKRPVY	360						
QY	361	PLQYDKDAEYGSNLKLPSTATAVEKKGSIYCOPELLCGEIELPRVNDYNGKKRYYY	420						
Db	361	PLQYDKDAEYGSNLKLPSTATAVEKKGSIYCOPELLCGEIELPRVNDYNGKKRYYY	420						
QY	421	ATEVQNSPPTAKIANKVGTQKVLHMGSDCHWSEPELFVSPAPAREDESVYTCVYSE	480						
Db	421	ATEVQNSPPTAKIANKVGTQKVLHMGSDCHWSEPELFVSPAPAREDESVYTCVYSE	480						
QY	481	PNKAPFLILLDAKTFKELGRATVNEHMLDHCMEIFQNDLGAFTE	526						
Db	481	PNKAPFLILLDAKTFKELGRATVNEHMLDHCMEIFQNDLGAFTE	526						
RESULT 2									
QYQJIS6	10	PRELIMINARY:	PRT:	566	AA.				
AC	QJQJIS6:								
PT	01-OCT-2000 (TREMblrel_15, Created)								
PT	01-MAR-2001 (TREMblrel_16, last sequence update)								
PT	01-JUN-2002 (TREMblrel_21, last annotation update)								
DB	beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).								
GN	BCT001 OR BCT00 OR BCT00.								
GN	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Gastera; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxID=10090;								
RN	111								
RP	SEQUENCE FROM N.A.								
RA	SPRAIN-057RL; TISSUE=KIDNEY;								
RA	WYSS A., WILTZ G.M., WEGON M.D., BRUGGER R., WYSS M., FILELLEIN A.,								
RA	BACHMANN H., HUNZIKER W.;								
RT	*Expression pattern and localization of beta, beta carotene 15,15'-								
RT	dioxygenase in different tissues.*;								
RL	Submitted (DEC-2000) to the EMBL/GenBank/CCPR databases								
RL	121								
RP	SEQUENCE FROM N.A.								
RA	SPRAIN-057RL;								
RA	Reichardt T.M., Gentleman S., Duncan T., Yu S., Wiggett B., Gantt F.,								
RA	Cunningham F.X.;								
RT	*Identification, expression and substrate specificity of a mammalian								
RT	beta-carotene 15,15'-dioxygenase.*;								
RL	J Biol. Chem. 0-0-0(2001)								
DR	EMBL: AJ278064; CAB92531.2;								
DR	EMBL: AF271981; AAC31982.1;								
DR	MBD: MG11925923; Bct001.								
DR	InterPro: IPRO04294; PF065.								
DR	Pfam: PF03055; RPE65_1.								

Seq	Dioxigenase: Oxidoreductase	Score	Length	666
Query Match	75.28; Score 2124; Id 11; Length 666;			
Best Local Similarity	71.28; Prod. No. 9, 40, 150;			
Matches 374; Conservative	85; Mismatches 64; Indels 2; Gaps 2;			
OY	1 METPKNKEHEPEPKAEVQGLPTWVQGLNKNKGEMHTIGCTKNNHFGALALISF	60		
DB	1 MEITFQNNKKFOLEPQAVQVTSIPANIGLTLNKNKGMHTGVDSKYNHIFKRLALISF	60		
OY	61 TFRNGFVYVYSKYLPSDTYVNLNANLVSEPTMAVPPQPKNIPAKAFSTLSHT	120		
DB	61 SIQDEGFVYSKYLQSDTYIANIPNNIVSEPTMAVPPQPKNIPAKAFSTLSHT	120		
OY	121 TDNCLINIKMTGDDYATSETFIKKIDPQTLFLDWDVDSKVAVNLAISHHRYGACN	180		
DB	121 TDNCLINIKMGCGDFPATTEINIKKIDPQTLLELVDRKKVAVNLAISHHRYGACN	180		
OY	181 ILNNGSTVDCGTRKIVLEKIPSSVPEKEK-KSCFKELEVCSIPSRSLQSYVHSFG	240		
DB	181 VLNNGSTVDCGTRKIVLEKIPATVPDSKKKKSGVNAHAFVCSISRSRLSYVHSFG	240		
OY	240 ITENYVLEIGCPKRLDYLALVLCQVNMASCSLSEFPEYTHVHGVPTKFEVSTKY	299		
DB	241 VTEHYVLELPQPKLIDLMATVAVKRGVSMASCSLSEFDEKTYIHIIIDTKKPVTKKY	400		
OY	300 TDALVLYVHNINAVEEDGHVFDIVAVKNSLVDMFYLKLLDKQFVNNKLTSTPTGRFV	459		
DB	301 TDENVYVHNINAVEEDGCVLFDIVAYEEDSLYQLFLYANLNKDEEKSRLSTVPLRFA	460		
OY	360 VPLQYVKNVAVNSLNFV-PTSAVTPKSGSTVGGHITETSETPAVVYVNRFFKY	418		
DB	361 VPLVIRVNAVNSLNVKVSALANLNDSHVVGVTVLPHSLTPRLINAVNINRFFKY	420		
OY	419 VYAVEVMSDPTKTLKNVGRFVYVHNGEYVWPSPEIPVSDVAEENH-VGLTAVVY	478		
DB	421 IFAAEVWVSWPTKILKXDLITKSLKSESCWMAFLFVPTPGAKDEHGVTLTAVS	480		
OY	479 SEFNKAPFLIIDAKTFKFGATVYVNSMHLDEGMEIPONDGAA 523			
DB	481 TDQOKLPLFLIDAKSTFLAAVSDALMHLDEGLTPIPADNNA 525			
RESULT 3				
Q9ERN9	PRELIMINARY; PRI: 556 AA.			
AC	Q9ERN9			
DT	01-MAR-2001 (TrEMBLrel. 15, Created)			
ED	01-MAR-2001 (TrEMBLrel. 15, last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, last annotation update)			
DE	beta, beta-carotene 15,15'-dioxigenase (EC 1.13.11.21).			
OS	Mus musculus (Mouse).			
OC	Faryngata, Mollusca, Chordata: Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia, Eulipidia, Eudelphia, Scirodanthi, Muridae; Mus.			
OX	NCBI_TaxID=10090;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
PC	SPVAIN-CTSL; Tissue=KIDNEY;			
FA	Yan W., Jiang G.-F., Haeseleer F., Esnuli N., Chan J., Kertman M.,			
RA	Chowdhary G., Cammerhoffer P., Polyzoukakis, J., Zeng D.J.			
RT	"Cloning and characterization of a human beta, beta-carotene 15,15'-			
PT	dioxygenase that is highly expressed in the retinal pigment			
RI	epithelium."			
FI	5-Limit4 (Aug. 2000) from the EMBL database.			
DR	EMBL: AF294899; A01591.1;			
DR	MGD: MGI:1926923; Bcd01.			
DR	InterPro: IPR034294, IPR05.			
DR	Pfam: PF03055; RPE65_1.			
DR	Dioxygenase; Oxidoreductase.			
SV	SEQUENCE 556 AA; 6.852 KM; 20439/9180450700 Ck664;			


```

10b 121 IDNGLINIMKGEFFATSTNTYIPKINPOTLETTIKVYPRKVAANALSHPHDEAN 160
10b 181 ILNMGTSIVNGKRTKRYIPKIPSSVP-KEKKSCKPHLKVCSIPSKLADPSYHSG 239
10b 181 VLNMGTSIVNGKRTKRYIPKIPSSVP-KEKKSCKPHLKVCSIPSKLADPSYHSG 240
10b 240 ITENVVTEFLQPPFLDILAKMAVTPMMSASLAFHFKTKYTHLLCPVJFHY 300
10b 241 VTEENVTEFLQPPFLDILAKMAVTPMMSASLAFHFKTKYTHLLCPVJFHY 300
10b 300 TDAVVEFHNAVEEDGIVEFVIAVEDNSLYOFPIANINQPFKNSKLSVPLRFA 360
10b 301 TDAVVEFHNAVEEDGIVEFVIAVEDNSLYOFPIANINQPFKNSKLSVPLRFA 360
10b 360 VPLQYIDAFVGSNLVYL-PTSAFAVEKKSSTYCCPPTLCPGTELPVNTYKRY 418
10b 361 VPLQYIDAFVGSNLVYL-PTSAFAVEKKSSTYCCPPTLCPGTELPVNTYKRY 420
10b 419 VYATEVQMSVPKRIAKLNVQTEKVLHMGDHCMPSEPIFVSPDAREDEGVITGVV 478
10b 421 VFATGVQMSVPKRIAKLNVQTEKVLHMGDHCMPSEPIFVSPDAREDEGVITGVV 480
10b 479 SEPKAPFLDILAKTEKELGRATVNVEMHLDHGMFTFOND 520
10b 481 TDPOKLPFLDILAKSTELAKASVVDVMDHDLHGLFTDMD 522

```

RESULT 6

```

10b 090WH5 PRELIMINARY: PRT: 547 AA
10b 01-OCT-2000 (TREMURel. 15, Created)
10b 01-OCT-2000 (TREMURel. 15, Last sequence update)
10b 01-JUN-2002 (TREMURel. 21, Last annotation update)
10b CDNA FLJ10730.115, clone NT2RP1001107
10b Homo sapiens (Human)
10b Eukaryota, Metazoa, Chordata, Cranialia, Vertebrata, Euteleostomi,
10b Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo,
10b NCBI_TaxId=9606,
10b 11
10b SEQUENCE FROM N.A.
10b Isotal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
10b Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
10b Watanabe M., Hosokawa T., Kato Y., Kodaira H., Sogawa M.,
10b Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
10b Yamamoto T., Wakamatsu A., Nakamura Y., Nishihara K., Masuko Y.,
10b Nishimura Y., Iwaguchi T.,
10b *NEO human cDNA sequencing project.
10b Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases
10b EMBL: AK001592; BAA91776.1;
10b InterPro: IPR004294; RPF65;
10b Pfam: PF04055; RPF65;
10b SEQUENCE 547 AA: 62579 MW: 194418016660EF CP664;

```

Query Match 74.68; Score 2107; DB 4; Length 547;

Best Local Similarity 70.38; Pred. No. 1,6e-148;

Matches 367; Conservative 96; Mismatches 57; Indels 2; Gaps 2;

```

10b 1 METIFNKNKEHPEPIKAEOYQOLPTWLOGVLLRNPGMHTIGDKRYNMFGALHLSF 60
10b 1 METIFNKNKEHPEPIKAEOYQOLPTWLOGVLLRNPGMHTIGDKRYNMFGALHLSF 60
10b 1 MDIIFGNKRKEQLEPRKAVYQKIPAMLOGTILRNPGMHTIGDKRYNMFGALHLSF 60
10b 61 TIFNKEEYYSKYLRSYNTNINENRIVVSEGTWAYDPCKNIFSKAFSYLSHTIDF 120
10b 61 TIFNKEEYYSKYLRSYNTNINENRIVVSEGTWAYDPCKNIFSKAFSYLSHTIDF 120
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 181 ILNMGTSIVNGKRTKRYIPKIPSSVP-KEKKSCKPHLKVCSIPSKLADPSYHSG 239

```

```

10b 181 VLNMGTSIVNGKRTKRYIPKIPSSVP-KEKKSCKPHLKVCSIPSKLADPSYHSG 240
10b 240 ITENVVTEFLQPPFLDILAKMAVTPMMSASLAFHFKTKYTHLLCPVJFHY 300
10b 241 VTEENVTEFLQPPFLDILAKMAVTPMMSASLAFHFKTKYTHLLCPVJFHY 300
10b 300 TDAVVEFHNAVEEDGIVEFVIAVEDNSLYOFPIANINQPFKNSKLSVPLRFA 360
10b 301 TDAVVEFHNAVEEDGIVEFVIAVEDNSLYOFPIANINQPFKNSKLSVPLRFA 360
10b 360 VPLQYIDAFVGSNLVYL-PTSAFAVEKKSSTYCCPPTLCPGTELPVNTYKRY 418
10b 361 VPLQYIDAFVGSNLVYL-PTSAFAVEKKSSTYCCPPTLCPGTELPVNTYKRY 420
10b 419 VYATEVQMSVPKRIAKLNVQTEKVLHMGDHCMPSEPIFVSPDAREDEGVITGVV 478
10b 421 VFATGVQMSVPKRIAKLNVQTEKVLHMGDHCMPSEPIFVSPDAREDEGVITGVV 480
10b 479 SEPKAPFLDILAKTEKELGRATVNVEMHLDHGMFTFOND 520
10b 481 TDPOKLPFLDILAKSTELAKASVVDVMDHDLHGLFTDMD 522

```

RESULT 7

```

10b 090WH4 PRELIMINARY: PRT: 516 AA.
10b 01-DEC-2001 (TREMURel. 19, Created)
10b 01-DEC-2001 (TREMURel. 19, Last sequence update)
10b 01-MAR-2002 (TREMURel. 20, Last annotation update)
10b Putative b,b'-carotene-15,15'-dioxygenase.
10b B-DIOX.
10b Bradydactylus (Zebrafish) (Zebra danio).
10b Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi,
10b Actinopterygii, Neopterygii, Teleostei, Ostariophysi, Cypriniformes,
10b Cyprinidae, Danio.
10b NCBI_TaxId=7955,
10b 11
10b SEQUENCE FROM N.A.
10b TISSUE-LIVER.
10b Kiefer C., Hessel S., Lampert J.M., Vogt K., Lederer M.O.,
10b Brechtel D.E., vonLintig J.,
10b *Identification and characterization of a Mammalian enzyme catalyzing
10b the asymmetric oxidative cleavage of provitamin A.
10b J. Biol. Chem. 276:14110-14116(2001).
10b EMBL: AJ293300; GAC37566.1;
10b InterPro: IPR004294; RPF65;
10b Pfam: PF04055; RPF65;
10b DIOXENASE.
10b SEQUENCE 516 AA: 58584 MW: 100447707655794F CP664;

```

Query Match 60.78; Score 1715; DB 14; Length 516;

Best Local Similarity 59.78; Pred. No. 2.0e-119;

Matches 310; Conservative 87; Mismatches 114; Indels 8; Gaps 4;

```

10b 1 METIFNKNKEHPEPIKAEOYQOLPTWLOGVLLRNPGMHTIGDKRYNMFGALHLSF 60
10b 1 METIFNKNKEHPEPIKAEOYQOLPTWLOGVLLRNPGMHTIGDKRYNMFGALHLSF 60
10b 1 MOYDGNKKEHPEPIKAEOYQOLPTWLOGVLLRNPGMHTIGDKRYNMFGALHLSF 60
10b 61 TIFNKEEYYSKYLRSYNTNINENRIVVSEGTWAYDPCKNIFSKAFSYLSHTIDF 120
10b 61 TIFNKEEYYSKYLRSYNTNINENRIVVSEGTWAYDPCKNIFSKAFSYLSHTIDF 120
10b 61 AINKEVIVSPKLSKISYNSNCAPIVAVSEGTWAYDPCKNIFSKAFSYLSHTIDF 120
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 121 TDNGLINIMKTGDYATSETNIRKIDPOTLETLAKVYSKYVAVNATSHPHDSAGN 180
10b 181 ILNMGTSIVNGKRTKRYIPKIPSSVP-KEKKSCKPHLKVCSIPSKLADPSYHSG 239
10b 181 ILNMGTSIVNGKRTKRYIPKIPSSVP-KEKKSCKPHLKVCSIPSKLADPSYHSG 239

```

[illegible]

```

Query Match 39.5% Score 115.5/754 Length 545;
Best Local Similarity 45.5% Recd 9.2e-75;
Matches 233 Conservative 96; Mismatches 19; Gaps 9

0Y 10 EEEHPKAEVOCGLPTWLGVLVLANGGMHITGDTKNWEDGIALHSPTEKGEVY 69
Fb EEAFFSLASAWWHFFRWMLNGSLFLCKEKEFEKCKINHFELCMALHFFMAFCHVY 87
0Y 28 EAAFFSLASAWWHFFRWMLNGSLFLCKEKEFEKCKINHFELCMALHFFMAFCHVY 87
0Y 70 PSATYSTPTNYNENENFVSESEPTMAFPCDPRITAKAAS YSEHTFEETGCLIN 127
Db KSTEDLSADYIKANSAANKAVISEPILADPEKCNVFEELMSKEFLGNAAMATGMINA 147
0Y 128 IMKIQVYAASTFENIKKIPQPTFTIKPKYNSKAAVAVNATSHRPHDSAGNINMST 187
Lb IVYKRIKLYLCELETRNMKNVLELELELEENWELWELVANSALAHNHYGLGSLAHM 207
0Y 188 IVDKGRTKYVLEKIPSSVEPEKKEKSKCKLEWVGSLSHSLLDYSYNSGCTENYV 247
Db PGDYG-PEYKVPVIVD---PEKVAQETIHGAVAVTSLASIEKQKSYNSHPCPTVYIF 263
0Y 242 IEEPEFLDVKLAIAVIGVNMVNSLSHSHEKCTKMEHEVPEKKEFEVSTFETVALV 307
Lb IELPIKMLKRLKIALISLIPKAFSLSLNPEFLWNIHFIVVKKPLDGLIPYVKKIPV 423
0Y 308 HINAVEGHVVEFDIVAVPENSLSYDFYKILAKL-----DFEVNNRTSITPKRPV 362
Lb GIMAHNPGGVVITDQTEPHETFLPVYCTQMLPACDEGLFVIRVAAKCTET--RREVL 381
0Y 363 QYKELAEVNSNLKPLTSATAVKEKESGTCGPEILCE -- GTELEKVNITVMS 413
Db NVSLNAPEDGNISPLSYLSAVNAAKAVATINWSEHNHQLDELEKQGLDEPGLVYIRP 441

```

414 KKKYVATVGVMSVPTKIAKINVOCTREKLEHMGEDHCEHSPFVPSDAREDEGVVL 473
 442 KKHFFFGGGR-HLVGDSLLKVDVYNKTLKVKREGCTPSEPVVAPAGTNEEDGVVL 500
 474 TCYVVSPPKAPPLLLDAKTEKELGRATVNMHLDGMETP 517
 501 SVVITPNQNESNPLVLDAKNPEELGRAVPMVGMPCGHTETP 544

RESULT 10

099NFI PRELIMINARY: PRT: 532 AA.
 ID 099NFI
 AC 099NFI
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE B.D-carotene 9,10'-dioxygenase.
 GN BCD02 OR B-DIOX-11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=21216714; PubMed=11278918;
 RA Kleier G., Hessel S., Lampert J.M., Vogt K., Lederer M.O.,
 RA Breitaupt D.E., von Linling J.,
 RT "Identification and characterization of a mammalian enzyme catalyzing
 RT the asymmetric oxidative cleavage of provitamin A."
 RL J. Biol. Chem. 276:14110-14116(2001).
 DR EMBL: AJ290492; F062825; 1.
 DR MGI: MGI:2177469; Bcd02.
 DR InterPro: IPR004294, RPE05.
 DR Pfam: PF03055; RPE05; 1.
 KW dioxygenase.
 SO SEQUENCE 532 AA; 60142 MW; 7461AD5A54FE6B43 CRC64;

Query Match 39.48; Score 1112.5; DB 11; Length 532;
 Best Local Similarity 43.58; Pred No 1.5e 74;
 Matches 228; Conservative 94; Mismatches 175; Indels 27; Gaps 10;

10 EEHPEPIKAEVQGLPTWLOGVLLRNGPQMHTIGTKYNNHMDGLALLHSTFFKNGVY 69
 19 EEFLSAVASARVGRHPEHMGVLLRVGRKPEFGKQRYNHWDGALLHGFEMRGVTV 78
 70 RSKYLSLDVYNCNIEANRTIVSHFGTMADPCKNIFAKAFSYLSHTLPEDNCLNIM 129
 79 KSKFLQSDTYKANSAGRIIVISFGTLALDPCKSTFERFKSRFP--PTMTDNTNVAV 136
 130 KTDIDVYATSEINFIRKIDPQLETLKVDYSKYVAVNLATISHPHYSAGNLLNGTSIV 189
 137 QYGGDYVMSTETNFMKVDIEMLERFEKVDMSKFLAVNGATAHPRYDGTAYNMGNSYG 196
 190 DKORTYVLEKLISSVPEKEKKSKCKHLEVCSSIPSRSLDQSYHSGITENTIVTE 249
 197 PRG-SYNNITFPV---PKRKEPETIHAGVLCSTASTEKMKRSTYSHSGMKNTLIVE 252
 250 QPKRIDIVKLATAYINGVNASCLSFHKEDKTWPFVDRKTKKEVSTFYDALVYHHI 309
 253 QPKMKIMKILTSKIGKPRFADGISERPQYNTFRHVVDKHTQLLRPGMYSPRLTNOI 312
 310 NAVEHGHVVEFDIVAKRNSLDYMFYLR-----LDKDFVNNKLTISPRKRFVPL 362
 313 NAVEHGGCIIVIDCCDGRSLDIYOLQNIKRAGRCTQVYEL--KANSFP--PRFVPL 368
 363 QYKIDAEVGSNLVKLP--TSATAVKEKDSIFGPELCE-----GIELPRVND--YNG 413
 369 IWSVDAAEKKNLSPLSYSSASAVKQGTGFTWSPNNIHENI FREGDIEFPQDIYNGPFG 428
 414 KKKYVATVGVMSVPTKIAKINVOCTREKLEHMGEDHCEHSPFVPSDAREDEGVVL 473

Db 429 KKKYVATVGVMSVPTKIAKINVOCTREKLEHMGEDHCEHSPFVPSDAREDEGVVL 487
 QY 474 TCYVVSPPKAPPLLLDAKTEKELGRATVNMHLDGMETP 517
 Db 488 SVVITPNQNESNPLVLDAKNPEELGRAVPMVGMPCGHTETP 544

RESULT 11

096JYS PRELIMINARY: PRT: 539 AA.
 ID 096JYS
 AC 096JYS
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE CONA FLJ14895 fis. Clone JACR190449, weakly similar to B.taurus
 DE retinal pigment epithelial membrane receptor p63.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA.
 RA Isogai T., Gao F., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 PA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 PA Natamura Y., Nagahara K., Masuda Y., Sasaki N.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK827801; HA85479; 1.
 DR InterPro: IPR004294; RPE05.
 DR Pfam: PF03055; RPE05; 1.
 SW SEQUENCE 539 AA; 60846 MW; 6434612331B86E1A0C64;

Query Match 39.08; Score 1102.5; DB 4; Length 539;
 Best Local Similarity 44.38; Pred No 8.4e 74;
 Matches 232; Conservative 85; Mismatches 178; Indels 29; Gaps 10;

10 EEHPEPIKAEVQGLPTWLOGVLLRNGPQMHTIGTKYNNHMDGLALLHSTFFKNGVY 69
 28 EEMPEGLASVVMHPEHMGVLLRVGRKPEFGKQRYNHWDGALLHGFEMRGVTV 87
 70 RSKYLSLDVYNCNIEANRTIVSHFGTMADPCKNIFAKAFSYLSHTLPEDNCLNIM 127
 88 RSKFLQSDTYKANSAGRIIVISFGTLALDPCKSTFERFKSRFP--PTMTDNTNVAV 147
 128 IMFTDQYVATSEINFIRKIDPQLETLKVDYSKYVAVNLATISHPHYSAGNLLNGTS 187
 148 YVYKCDYVLEGTETNFMKVDIEMLERFEKVDMSKFLAVNGATAHPRYDGTAYNMGNS 207
 188 IVDKRTYVLEKLISSVPEKEKKSKCKHLEVCSSIPSRSLDQSYHSGITENTIVTE 247
 208 FQYGG-ESYKVLVPR---PEKVIDLHLLGVVGLTSLHSEKSKSYSHSGMTENTIV 264
 248 LEQPKRIDIVKLATAYINGVNASCLSFHKEDKTWPFVDRKTKKEVSTFYDALVYHHI 307
 264 LEQPKMKIMKILTSKIGKPRFADGISERPQYNTFRHVVDKHTQLLRPGMYSPRLTNOI 324
 308 HINAYEDGHVVEFDIVAKRNSLDYMFYLR-----DPEVNNKLTISPRKRFVPL 362
 324 QINAPFLQGSVLIIDCCDGRSLDIYOLQNIKRAGRCTQVYEL--KANSFP--PRFVPL 368
 363 QYKIDAEVGSNLVKLP--TSATAVKEKDSIFGPELCE-----GIELPRVND--YNG 413
 382 NWSLNPEDCNLSPLSYSSASAVKQGTGFTWSPNNIHENI FREGDIEFPQDIYNGPFG 441
 414 KKKYVATVGVMSVPTKIAKINVOCTREKLEHMGEDHCEHSPFVPSDAREDEGVVL 473
 442 KKHFFFGGGR-HLVGDSLLKVDVYNKTLKVKREGCTPSEPVVAPAGTNEEDGVVL 494
 474 TCYVVSPPKAPPLLLDAKTEKELGRATVNMHLDGMETP 517
 501 SVVITPNQNESNPLVLDAKNPEELGRAVPMVGMPCGHTETP 544

RESULT 12
 Q9BYV7 PRELIMINARY: PRT: 556 AA.
 ID Q9BYV7: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE Putative b,b'-carotene 9',10' dioxygenase.
 GN B-DIOX-II.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=21216714; PubMed=11278918;
 RA Kiefer C., Hessel S., Lampert T.M., Vogt K., Lederer M.O.,
 RA Breithaupt D.E., von Lintig J.;
 RT "Identification and characterization of a Mammalian enzyme catalyzing
 RT the asymmetric oxidative cleavage of provitamin A.";
 PL J. Biol. Chem. 276:14110-14116(2001)
 DP EMBL: AJ290393; CAC37904.1;
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 KW Dioxygenase.
 SQ SEQUENCE 556 AA: 62882 MW: 2082319628866P CRC64;
 Query Match 39.0%; Score 1102.5; DR 4; Length 556;
 Best Local Similarity 44.1%; Pctd No. 87; 74;
 Matches 231; Conservative 87; Mismatches 183; Indels 23; Gaps 9.

QY 10 EHEPERIAEVOGQLPTWLOGVLLRNGSMHTIGDKYNNHFDGLALHSFEKGEVY 69
 DB 39 BEAPGIGISARVWGHFPRKWLNGSLRLRIGKFEFGKSDKYNHFDGMLLHOFRAKGVY 98
 QY 70 RSKYLRPTDYNCFNANFIVSEPTGMAXRPPCKNIFAKFS--YLSHTIPEPTDNLIN 127
 DB 99 RSKFLOSOTYKANKNAKNIIVISERGTALRPPCKNVEPRNSPELFGKAAAMTDNNV 158
 QY 128 IMKTGDDYYATSENFIRKIDPOTLETLKDYKSKYVAANLATSHPHYSAGNLLNGTS 187
 DB 159 YKRYGDDYICTETFPNKKVDIETLEKTEKYSKFIYAVNGATNHPHYDDGTAYNNGS 218
 QY 168 IYDKRRTYVLEKIPSSVPEKSKKSKHLEVCISIPRSLLDPSITYHSFGITENTYV 247
 DB 219 FGPRVQ-FSKYVLRV--PEEVLDGETI HGQVVICISIASTEKGRPSYHSHFGMTNVI 274
 QY 248 IEOPKRLIVLATAYIRGVNMAASCLSFHKEDKTFHFVDRKTKKEVSTKFTYDALVYH 307
 DB 275 IEOPKRLIVLATAYIRGVNMAASCLSFHKEDKTFHFVDRKTKKEVSTKFTYDALVYH 307
 QY 308 HINAVEEDGSHVFDIVAYRNSLYDMFYLKLDK-----DEYNNKLTISIPCKKVVPL 362
 DB 335 QINAEEDGSHVFDIVAYRNSLYDMFYLKLDK-----DEYNNKLTISIPCKKVVPL 362
 QY 363 QYDKRAEAGSNLYKLP-TSATAVNEKDSIYCQPELCE-----GIELPKVNYD-YNG 413
 DB 392 QYDKRAEAGSNLYKLP-TSATAVNEKDSIYCQPELCE-----GIELPKVNYD-YNG 413
 QY 393 NVSINAPRPNISPISTASAVKQAPSTICSHFNH HDELEKEEDSIEFPQIYIDFSG 452
 DB 414 KRYKYVATEVQMSFVPTKIAKLNVQTEVYLHMGECDSSEPTFVSPSPAREDEGVYL 473
 QY 453 KRYHYFYCGGR-FLVGDSLIKVDVYKTLKVMREDGPRSEPRFVAPAGINHHGCVIL 511
 DB 474 TCYVVSSEPKAPRFLILDAKTFKELGRATVNVEMHLLHGMFTP 517
 QY 512 SVVLIIPNONSFLLVLDAKNEELGRAEVVVMYSGHGFIP 555

RESULT 13
 Q91Z05 PRELIMINARY: PRT: 533 AA.
 ID Q91Z05: 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE Putative b,b'-carotene 9',10' dioxygenase.
 GN B-DIOX-II.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=RETINA;
 RA Chang R., Hawes N.L., Hurd R.P., Davisson M.T., Heckenlively J.;
 RT "A point mutation in Rpe65 gene causes retinal degeneration (rd12) in
 RT mice.";
 PL Mice.
 DP FMRU: AF410461; AL011119.1;
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 SW SEQUENCE 533 AA: 61066 MW: 243555411442 CRC64;
 Query Match 39.1%; Score 1077; DR 11; Length 533;
 Best Local Similarity 42.2%; Pctd No. 6,56-73;
 Matches 222; Conservative 89; Mismatches 131; Indels 24; Gaps 12.

QY 2 ETEFNKKEHEPERIAEVOGQLPTWLOGVLLRNGSMHTIGDKYNNHFDGLALHSFE 61
 DB 17 ETV-----EELSPLTAHVTRIPRLMELTGLSGPGLFVSPSPFVH FVQALALHPRD 71
 QY 62 FVNDVYVYFETTYTNCNENIENRIVETNCTMAFTYVNTHTAFVGLSEHTIPEPT 121
 DB 72 FVNDVYVYFETTYTNCNENIENRIVETNCTMAFTYVNTHTAFVGLSEHTIPEPT 121
 QY 122 DNCLINIKTGDYYATSENFIRKIDPOTLETLKDYKSKYVAANLATSHPHYSAGN 181
 DB 130 DNALNIVYVGEDVYACETENFIRKIDPOTLETLKDYKSKYVAANLATSHPHYSAGN 189
 QY 182 LMGSTYVDKPTKVLFTIPSSVPEKSKKFTHELVYVSTPSTIGPSTYHSFGITENTYV 241
 DB 190 YNIGKCFKNTVAANIIRIP--PLKADIEDPINKSEVVOFPCSDKRFPSVSHSGILT 246
 QY 242 ENYVIFVQPKLIVKATAY-IPVNMASCLSFHKEDKTFHFVDRKTKKEVSTKFTYDALVYH 307
 DB 247 PNYVIFVETPKINIFFLSSMSLGMANYKCFESNSKSMVLEVALKRPKRYFNKRY 306
 QY 301 LALVLYHINAVEEIGSHVFDIVAYRQD-NSLYDMFYLKLDKDFEV--NNKLTISIPCK 356
 DB 307 SFENLHNNINYEEDGFLIVDLCKMGKGEFVYNYLYANIRFMLEVKKRMAKAPDQFVR 366
 QY 367 FVNVETVQKDAEAGSNLYKLP-TSATAVNEKDSIYCQPELCE-----GIELPKVNYD-YNG 410
 DB 367 FVNVETVQKDAEAGSNLYKLP-TSATAVNEKDSIYCQPELCE-----GIELPKVNYD-YNG 410
 QY 411 YNEKRYVYVATEVQMSFVPTKIAKLNVQTEVYLHMGECDSSEPTFVSPSPAREDEGVYL 470
 DB 426 FGKRYVYVATEVQMSFVPTKIAKLNVQTEVYLHMGECDSSEPTFVSPSPAREDEGVYL 470
 QY 471 VLTGVVVSSEPKAP-FLLILDAKTFKELGRATVNVEMHLLHGMFTP 515
 DB 485 VLVSVVSPGAGOKPAYLVLINAKDISEIAKAVEETNIIVTFPHGIF 530

RESULT 14
 Q16518 PRELIMINARY: PRT: 533 AA.
 ID Q16518: 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE Retinal pigment epithelium-specific 61 kDa protein.
 GN RPE65.

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CC NCBI_TaxID=9606;
 RN
 RP SCIENCE FROM N.A.
 RC TISSUE=RETINAL PIGMENT EPITHELIUM;
 RX MEDLINE=9535969, PubMed=763413,
 RA Nicoletti A., Wong D.J., Kawase K., Gibson L.H., Yang-Feng T.L.,
 RA Richards J.E., Thompson D.A.;
 RT "Molecular characterization of the human gene encoding an abundant 61
 RT kDa protein specific to the retinal pigment epithelium";
 RL Hum. Mol. Genet. 4:641-649(1995).
 RN 121
 RP SEQUENCE FROM N.A.
 RA Morimura H.;
 RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U18991; AAC99012.1; -;
 DR EMBL: U20510; AAC14586.1; -;
 DR EMBL: U20476; AAC14586.1; JOINED.
 DR EMBL: U20477; AAC14586.1; JOINED.
 DR EMBL: U20478; AAC14586.1; JOINED.
 DR EMBL: U20479; AAC14586.1; JOINED.
 DR EMBL: U20481; AAC14586.1; JOINED.
 DR EMBL: U20482; AAC14586.1; JOINED.
 DR EMBL: U20484; AAC14586.1; JOINED.
 DR EMBL: U20485; AAC14586.1; JOINED.
 DR EMBL: U20486; AAC14586.1; JOINED.
 DR EMBL: AF039868; AAC39660.1; -;
 DR EMBL: AF039865; AAC39660.1; JOINED.
 DR EMBL: AF039857; AAC39660.1; JOINED.
 DR EMBL: AF039858; AAC39660.1; JOINED.
 DR EMBL: AF039859; AAC39660.1; JOINED.
 DR EMBL: AF039860; AAC39660.1; JOINED.
 DR EMBL: AF039861; AAC39660.1; JOINED.
 DR EMBL: AF039862; AAC39660.1; JOINED.
 DR EMBL: AF039863; AAC39660.1; JOINED.
 DR EMBL: AF039864; AAC39660.1; JOINED.
 DR EMBL: AF039865; AAC39660.1; JOINED.
 DR EMBL: AF039866; AAC39660.1; JOINED.
 DR EMBL: AF039867; AAC39660.1; JOINED.
 DR InterPro: IPR004294; PPF65.
 DR Pfam: PF03055; PPF65.1
 SV SEQUENCE 533 AA; 60947 MW; 7193C93F3325798D CRC64;
 Query Match 38.1%; Score 1075; DB 4; Length 533;
 Best Local Similarity 42.2%; Pred. No. 9, 20-72,
 Matches 222; Conservative 93; Mismatches 187; Indels 24; Gaps 12;
 QY 2 EITFNRRKEHPERPIKAEOGOLPTWLOGLVLRNPGOMITGCTRYNNHMFGLALHSFT 61
 DB 17 ETV-----ELSTPLTAHAYGRIPLMLTSLRGCGLFEGVSEPPYHFLDQALHMFED 71
 QY 62 FRNGEYVRSKYLPSTTYNCNIPARIVSEHGMAYRPPCKNIJAKAFSYSHIIPFT 121
 DB 72 FRKHVTVYRPPPTPTDAVYPAAMTKRPIVITFERGTCAFRPPCKNITSPPFSYRQV--EVT 129
 QY 122 ENCLINIMKTGDYVATSETNFIKIDPOTLETLDKVDYSKYVAVALATSHPHYSAGNI 181
 DB 130 DNALVNVYRPGELGYACETNFTIKINPTLETIKQVLCNVSVNGATAIAPHINDGIV 189
 QY 182 LNMGTSLVKGRTKRYLFLIPSSVPEKEKKSCFKLELVGCIIPSKSLDSYHSFOIT 241
 DB 190 YNIGNCGKKNSTAYNIIVKIP--PLQADKEPDISKSEIIVQFPCSDRFPKPSYHSGILT 246
 QY 242 ENYIVFLEDFPKIDYKILATAY-IRGVNNAASCLSPHKEDKTPHFHVDRKTKKEVSTRKYT 300
 DB 247 PNYIVETVETVKINLEKFLSSWSLWGANYMDFESNENMGVYLHIAKKRKKRYLNNKRT 306
 QY 401 DALVLYHHINAVFEDCHVVDVAVRD-NSLYDMFYLKKLQDFE---VNNKLTSIPCK 356
 DB 307 SPNPLPHHINTYEDNGFLIVDLICQMKGFEEVYNYIYIAMLRENWEVKKNAKRAVQPEVR 366

QY 357 PFVVLGYCKDAEWSNIVLP-TSAFAVENKSYICOPELICG-----IEHPENVY 410
 DB 367 RYVPLINDK-ADTGKNLVTLPNTVTALVLSNETIWLLEPPVLSPPRQVAPFQYINVK 425
 QY 411 YNCKRYKRYVVAIEVWSVPEPIALINWVLECVIWMJELHWSESEIVSFAKRELE 470
 DB 426 YCKRYVYAVGDSI NHH-VPHGCKINVAIKFLWVQFHSVSEPIVSDALEEDQ 484
 QY 471 VVLGVVSEPNKAP-FLILIAKTEKELGKATVANNEMILDIKMF 515
 DB 485 VLVSVVSPGAGVPAVILILINAKRIDSEVAPARVETINIVPFGHF 540
 RESULT 15
 ID 070276 PRELIMINARY; PRT; 533 AA.
 AC 070276;
 DT 01-AUG-1998 (TREMURel. 07, Created)
 DT 01-AUG-1998 (TREMURel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMURel. 19, Last annotation update)
 DE Retinal pigment epithelium specific protein.
 GN RPE65.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=MISTAR; TISSUE=RETINA;
 RX MEDLINE=98171299; PubMed=9512345;
 RA Wang C., Tiedge P., Kucharsky J., Pages A., Schmitt-Bernard C.P.,
 RA Hanel C.P.;
 RT "Rat messenger RNA for the retinal pigment epithelium specific protein
 RT RPE65 gradually accumulates in two weeks from late embryonic days.";
 RL FEBS Lett. 423:133-137(1998).
 DR EMBL: AF035673; AAC40059.1; -;
 DR InterPro: IPR004294; PPF65.
 DR Pfam: PF03055; PPF65.1
 SV SEQUENCE 533 AA; 60989 MW; D80717646CF52C93 CRC64;
 Query Match 37.9%; Score 1072; DB 11; Length 533;
 Best Local Similarity 42.6%; Pred. No. 1, 50-71;
 Matches 224; Conservative 89; Mismatches 193; Indels 24; Gaps 12;
 QY 2 EITFNRRKEHPERPIKAEOGOLPTWLOGLVLRNPGOMITGCTRYNNHMFGLALHSFT 61
 DB 17 ETV-----ELSTPLTAHAYGRIPLMLTSLRGCGLFEGVSEPPYHFLDQALHMFED 71
 QY 62 FRNGEYVRSKYLPSTTYNCNIPARIVSEHGMAYRPPCKNIJAKAFSYSHIIPFT 121
 DB 72 FRKHVTVYRPPPTPTDAVYPAAMTKRPIVITFERGTCAFRPPCKNITSPPFSYRQV--EVT 129
 QY 122 ENCLINIMKTGDYVATSETNFIKIDPOTLETLDKVDYSKYVAVALATSHPHYSAGNI 181
 DB 130 DNALVNVYRPGELGYACETNFTIKINPTLETIKQVLCNVSVNGATAIAPHINDGIV 189
 QY 182 LNMGTSLVKGRTKRYLFLIPSSVPEKEKKSCFKLELVGCIIPSKSLDSYHSFOIT 241
 DB 190 YNIGNCGKKNSTAYNIIVKIP--PLQADKEPDISKSEIIVQFPCSDRFPKPSYHSGILT 246
 QY 242 ENYIVFLEDFPKIDYKILATAY-IRGVNNAASCLSPHKEDKTPHFHVDRKTKKEVSTRKYT 300
 DB 247 PNYIVETVETVKINLEKFLSSWSLWGANYMDFESNENMGVYLHIAKKRKKRYLNNKRT 306
 QY 357 PFVVLGYCKDAEWSNIVLP-TSAFAVENKSYICOPELICG-----IEHPENVY 410
 DB 367 RYVPLINDK-ADTGKNLVTLPNTVTALVLSNETIWLLEPPVLSPPRQVAPFQYINVK 425

OY 411 YNKKRYVAIEYQMSVP1K1AKLNVJKEVLHWGEJHCWPSPIFVPSPDAREDEG 470
DB 426 CGGKPYIYAAGLGNHF-VPUKCKLNVKIKELMMQEPUSPSEPIFVSQPUALHEDDG 484
OY 471 VVLTGVVVSHPNKAP-FLLTIDAKTFKEIGRATVNVEMHDLHGMF 515
DB 485 VVLSVVVSPGAGQKPAFLVLVNAKDLSEIARAETINIPVTFHGLF 530

Search completed: July 15, 2003, 09:29:40
Job time : 88 secs

RESULT 6
AAH14870
ID AAH14870 standard; cDNA: 1983 bp.
XX
AC AAH14870;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12717.
XX
KM Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000IP-0183767.
PR 09-JUN-2000; 2000IP-0241899.
XX
PA (HELIX) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
PS Claim 8; SEQ ID 12717; 2537pp + CD ROM, English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one or
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
CC AAB95883 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX
SQ Sequence 1983 BP; 583 A; 384 C; 493 G; 523 T; 0 other;

Query Match 8.88; Score 272.8; DB 22; Length 1983;
Best Local Similarity 51.18; Pred. No. 8.2e-57;
Matches 789; Conservative 0; Mismatches 722; Indels 33; Gaps 5;

QY 160 GAATGACATGCGACAGCTGATAAATGATGAGTGGAGAGTGCAGTGTGCTGCTGCA 219
DB 184 GAAAGAGCTTTCAGGAGGATCTCTGTCTAGTCTTGGGACATTTCTCAAGTGGCTCAAT 243

QY 229 GGGGACCTCTCTGCAAAATGGGCTGAGAGTATGATGAAATAGGAGGATGATATGATGAT 279
DB 244 GATTTTCTTATTTTAAATTTGACCTGGGAAATTTGATGTTGGGAGGATGATGATGATGAT 403
QY 280 TGGTTCAGTGGCTTGGCTTGGCTGAGTATGATGATGATGATGATGATGATGATGATGAT 439
DB 304 TGGTTTATGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 463
QY 340 ACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 499
DB 364 ACGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423
QY 400 GTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
DB 424 ATCTTGAATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
QY 454 AAGGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 513
DB 484 ATGTGCAAGCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
QY 514 ATATGAAAGAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
DB 544 TATGTCGAGTACAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
QY 574 GATCCAGAGCTCTGAGAGCTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 633
DB 604 GACATGAACTCTGGAAAAACAGAAAGATGATGATGATGATGATGATGATGATGATGATGAT 663
QY 634 TTGGCACTTCTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 693
DB 664 GGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 723
QY 694 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
DB 724 TTTT-----GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
QY 754 AAGGAAAGAGGAAATCTCTTTAAAGATGATGATGATGATGATGATGATGATGATGATGAT 813
DB 772 AAGGAGAGCTTGGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 831
QY 814 TCCCTCTGCAAGCAAGCTTACAGCAAGCTTACAGCAAGCTTACAGCAAGCTTACAGCAAG 873
DB 832 GAGAAAGAGGAAAGCTTCTTACAGCAAGCTTCTTACAGCAAGCTTCTTACAGCAAGCTT 891
QY 874 ATAGAGAGGAGGATTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
DB 892 ATGAGCAAGCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
QY 934 AAGCGGCTTCTGCT 993
DB 952 GCTTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
QY 994 AGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1053
DB 1012 AAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1071
QY 1054 CACATTAATGCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1104
DB 1072 CAAATCAATGCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1141
QY 1105 TACGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1164
DB 1132 GGAAGAGGCTTACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
QY 1165 GTGAGAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224
DB 1192 CAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1261
QY 1225 GACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1281
DB 1252 AGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1311

634 TTGGCACTTCACACCACTATGACAGTGTGCAAAATATTCACATGGGACTTCA 693
 694 ATTTGTGATTAAGGACAAATATGCTCTTAAAGATCCCTCTGCTGACAGAA 753
 695 GCAATCACTTATATTTATATATATATATATATATATATATATATATATAT 654
 696 TTTT-----GAGCAATATGCTCTTAAAGATATGCTGCTGCTGCTGCTGCT 655
 754 AAGCAAAAGAAATATCTTTTAAACACTGTAATGCTGCTGCTGCTGCTGCTG 813
 755 GAGGAGACCTTGGGAGAAATATATATATATATATATATATATATATATAT 703
 756 TTTT-----GAGCAATATGCTCTTAAAGATATGCTGCTGCTGCTGCTGCT 704
 814 TCCCTGCTCAACCAAGCTTACTACACAGCTTTGGAATCAACAAATATATGCT 873
 815 GAGAAAGGAAACCTTTCTTACTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 822
 874 ATAGAGCAACCTTTAACTGATATGCTCAACTGCAACTGCTGCTGCTGCTGCT 933
 875 ATTTAACCACTTAAATATATATATATATATATATATATATATATATATAT 823
 934 AACTGAGCTTCT 933
 935 GCTTTTACATGAT 942
 946 ACAAAGACGAAAGAAAGATATATATATATATATATATATATATATATATAT 1053
 947 AAGGCACTGCTGCT 1002
 1054 CACATTAATCT 1104
 1055 TACACAGCAATATCT 1164
 1056 GCAAGCAACCTTAGAAGCTTACCACTTACCACTTACCACTTACCACTTACCACT 1122
 1165 GTGAAGCAACAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1224
 1166 GAGTCT 1182
 1225 GACAGCAATCT 1281
 1226 AGTTGCAATGCT 1242
 1282 GTAAAGCAAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
 1283 GTGAAGCAACCTTAGAAGCTTACCACTTACCACTTACCACTTACCACTTACCACT 1302
 1330 GAGTCT 1380
 1331 GAAAGCAACCTTAGAAGCTTACCACTTACCACTTACCACTTACCACTTACCACT 1442
 1382 AAGTCT 1440
 1383 TATCATCT 1442
 1441 AATCT 1500
 1442 CATCT 1482
 1501 AATCT 1500
 1483 GTTTTCT 1542
 1541 GTTCT 1620
 1542 AATCT 1602
 1621 AATCT 1680
 1603 GAGTCT 1682

QY 1681 CC 1682
 DB 1663 CC 1664
 RESULT 8
 AAF79560
 ID AAF79560 standard; CDNA: 1724 BP.
 XX
 AC AAF79560;
 DE 29-MAY-2001 (first entry)
 XX
 DE Canine RPE65 CDNA.
 XX
 KW Canine, dog, RPE65, canine retinal pigment epithelium 65; RPE65;
 XX congenital stationary night blindness; CSNB; mutation detection; ss.
 XX
 OS Canis familiaris.
 XX
 PN US620114-B1.
 PD 13-MAR-2001.
 XX
 PF 30-AUG-1999; 99US-0385259.
 XX
 PR 06-OCT-1998; 98US-0103219.
 XX
 PA (CORR.) CORNELL RES FOUND INC.
 XX
 PI Aguirre GD, Acland GM, Ray K;
 DR WPI: 2001-265168/27.
 DR P-PSDB: AAB73955.
 PT Novel nucleic acid molecule encoding canine retinal pigment epithelium
 PT 65, where presence of mutation in one or both alleles is indicative of
 PT blindness.
 PT
 PS
 PS Claim 1; Fig 1A-1B; 25pp; English.
 CC The present sequence encodes canine retinal pigment epithelium 65 (RPE65)
 CC which has a sequence of 519 amino acids defined in the specification.
 CC The gene is useful for identifying dogs which are genetically
 CC normal, or are carriers of, or affected with congenital stationary
 CC night blindness (CSNB). Carriers for CSNB are identified through the
 CC detection of an AAG deletion at nucleotide 487-490 in the
 CC canine RPE65 gene. This allows a breeder to eliminate the carrier from
 CC the breeding stock or to breed carriers with genetically normal dogs.
 CC
 SQ Sequence 1724 BP; 465 A; 421 C; 294 G; 444 T; 0 other;
 Query Match 8.5%; Score 264.8; DB 22; Length 1724;
 Best Local Similarity 52.1%; Prod. No. 7.2e-55;
 Matches 81; Conservative 0; Mismatches 702; Indels 45; Gaps 8;
 QY 153 AAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 212
 DB 78 AACCGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 137
 QY 213 GCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 272
 DB 138 GCTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 197
 QY 273 GAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 342
 DB 198 TATCATCT 257
 QY 343 TATCATCT 392
 DB 258 GAGTCT 417

XX	oil seed/palm, chick pea, diagnostic, therapeutic, library, zebrafish,
KW	retinoid/vitamin A deficiency, beta-diox II, transgenic, nutrition,
KM	carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop; ss.
XX	
OS	Brachydanio rerio.
XX	
XX	Key Location/Qualifiers
FT	29..1678
FT	CD5
FT	/*tag= a
FT	/product= "zebra-2"
XX	/note= "Beta-carotene dioxygenase (beta-diox II)"
PN	
XX	WU200148163-A2.
PD	
XX	05-JUL-2001.
PF	
XX	27-DEC-2000; 2000NO-EP13273.
XX	
PR	24-DEC-1999; 99EP-0125895.
FR	20 MAR-2000; 2000EP 0105822.
XX	
PA	(GREE-) GENOVATION PFLANZENPROTEOMIK-NGF-SMPH
PI	
PI	Von Lintig J, Vogt K:
XX	
DR	WPI: 2001-425657/45
DR	P-PSDB: AAU04293.
XX	
PT	Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
PT	(beta-diox II) protein that cleaves beta carotene and lycopene to yield
PT	beta-apo-carotenol and beta ionone, and apolycoprenals, respectively
XX	
PS	Claim 8; Page 106-109; 116pp; English.
XX	
XX	The sequence represents the coding sequence of zebra-2, beta-carotene
CC	dioxygenase (beta-diox II). Beta-diox specifically cleaves beta-carotene
CC	and lycopene to form beta-apo-carotenol and beta-ionone, and
CC	apolycoprenals, respectively. The DNA is useful for transforming grain
CC	seeds, e.g. corn; oil seeds, e.g. Brassica seeds, edible seeds e.g. palm;
CC	edible seeds or seeds with edible parts e.g. chick-peas; potatoes,
CC	carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
CC	for diagnostic and/or therapeutic purposes and for preparing antisense or
CC	ribozyme type therapeutic agents and for detecting any abnormality of
CC	endogenous beta-diox II. The beta-diox II specific antisense
CC	oligonucleotides derived from the DNA sequence are useful for dose
CC	response studies in relevant models of retinoid/vitamin A deficiency
CC	during any stage of an organism's development. The nucleic acids are also
CC	useful as probes and as a guideline to define new PCR (polymerase chain
CC	reaction) primers for the cloning of substantially homologous DNA
CC	sequences from other sources. The nucleic acids are also useful for
CC	determining the presence or quantity of beta-diox II nucleic acid and
CC	determining presence and amount of beta-diox II. The polypeptide is also
CC	useful for increasing or decreasing the amount of beta-diox II levels in
CC	a cell or tissue which can modify the level of vitamin A and other
CC	retinoids. Antibodies are useful for studying beta-diox II localisation,
CC	screening of an expression library to identify nucleic acids encoding
CC	beta-diox II or the structure of functional domains. The transgenic
CC	fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
CC	organisms have improved nutritional quality or physiological condition
CC	as vitamin A aldehyde and retinoic acid, beta carotene or take up beta-
CC	carotene from the medium. Expression systems encoding beta-diox II are
CC	useful in the study of beta-diox II activity. Identification of cDNAs
CC	encoding beta-diox I and II allows the physiological characterisation of
CC	mammalian carotene/retinoid metabolism. Vitamin A production in crops and
CC	microorganisms can be achieved by transforming the organisms or crops
CC	with the cDNAs.
XX	
SQ	Sequence 2134 bp; 625 A; 448 C; 441 G; 620 T; 0 other,
Query Match	8.1%; Score 250.8; DB 22; Length 2134;
Best Local Similarity	50.8%; Prod No 2 2e-51;
Matches 789; Conservative	0; Mismatches 712; Indels 51; Gaps 6

[illegible]

RESULT 11
AAQ53070
10 AAQ53070 standard; cDNA to mRNA; 1599 BP

AC AAQ530703

IDT 02-JUN-1994 (first entry)
XY

Human retinol binding protein receptor cDNA

KW Retinol binding protein receptor; retinitis pigmentosa; ss
xx

Homosapiens

PN W09323538 A.

PL) 25 - NOV - 1993.

PP 14-MAY-1993; 03W0-1150045RR

15-MAY-1992; 9205-0883539.

PA (LUDW-) LUDWIG INST CANCER RES

PI BAVIK CO, ERIKSSON U, SIMON A,

WP1; 1993-386570/48

XX
XX

acid molecule - (used)

conditions and treat

PS Claim 1; Page 21-24; 44pp; English

(c) This sequence may optionally have a nucleotide extension at its

5'-region. The possible extensions are given in AA053071-73.

50 Sequence 1599 bp; 458 A; 354 C; 341 G; 446 T; 0 other

Query Match	7.98;	Score 245.4;	DB 14;	Length 1599;
Best Local Similarity	50.98;	Pred. No. 4.1e-50;		
Matches 790;	Conservative	0;	Mismatches 716;	Indels 45;
				Gaps 7

[illegible]

QY 1174 AAGCTTAC-----TCTATCTTAACTTCTTCTGCTCTCTAGAT 1224
 DB 1064 AAAAATCTAGAAAGCTCTCTTAACTTAACTTAACTTAACTTAACTT 1122
 QY 1225 GACAACTAGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 1244
 DB 1123 GATTAAGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 1142
 QY 1285 AAGAAAGAAAGATGAGAGATCTATTGTCACCTGAAATATTATGAGGG----- 1335
 DB 1183 CTGTGAGAGAGAGAGATCTATTGTCACCTGAAATATTATGAGGG----- 1242
 QY 1336 ---ATATTAAGCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1392
 DB 1243 GCAATTTGAGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1302
 QY 1393 GCAAGCAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1452
 DB 1303 TATTAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1462
 QY 1453 AAGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1512
 DB 1363 AAGAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1422
 QY 1513 AAGAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1572
 DB 1423 CACCCAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1482
 QY 1573 CCAA---ATAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1629
 DB 1483 GCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1542
 QY 1630 CGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1680
 DB 1543 AGGCTGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1593
 RESULT 12
 ID AAH07608 standard: cDNA; 825 bp
 AC AAH07608:
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SFG ID No.4443.
 XX
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss
 XX
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PE 28-JUL-2000; 2000EP 0116126
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241844
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -
 XX
 XX Claim 1: SFG ID 4443; 2537bp + CD ROM; English.
 PS
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises (a) an oligo dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH07608 to AAH1628 and
 CC AAH1633 to AAH1874; represent human cDNA sequences; AAH2445 to
 CC AAH2693 represent human amino acid sequences; and AAH1629 to AAH1632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 825 BP: 238 A; 159 C; 216 G; 204 T; 8 other;
 SO

Query Match 4.7% Score 145.6; DB 22; Length 825;
 Post local similarity 54.1%; Pred No. 1-25;
 Matches 286, Conservative 0, Mismatches 191, Indels 7, Gaps 2.

QY 160 GAAAGCT 219
 DB 184 GAAAGCT 243
 QY 220 GAAAGCT 279
 DB 244 GAAAGCT 303
 QY 280 GAAAGCT 339
 DB 340 GAAAGCT 399
 QY 464 GAAAGCT 423
 DB 400 GAAAGCT 453
 DB 424 GAAAGCT 483
 QY 454 GAAAGCT 513
 DB 484 GAAAGCT 543
 QY 514 GAAAGCT 573
 DB 544 GAAAGCT 603
 QY 574 GAAAGCT 636
 DB 604 GAAAGCT 667
 QY 633 CTG 636
 DB 664 ATGG 667
 RESULT 13
 ID AAH07608 standard: cDNA; 1265 bp.

[illegible]

Oy	965 GTTTTCCTTACGAGATATTAAGAAGCAGATAGTGATAAATAAAATTATTCGGGTCTCCG
Db	1028 AACTGATTTTAGAGATATTAAGATATCATATTTCAATTAATAATATGAGTATTAAGTAG
Oy	1012 GATGCACCAAGCTTTTACACGATGCGTGTGGCTGTATTAACCAACAATAAGCTTACAGAA
Db	1088 GTGCAACACTTACGAAT---CGAAGACCTTCTTCTACTGTAGCAATCATCTCAACTGCTTTGAA
Oy	1072 GAAGATGATGATGCTTTGTTTTTATTAATGCTGATGCTTACAG
Db	1145 CGGATGCGCACGCTGCTGTGATGACATATTCAGACTACAG
RESULT 15	
AAS07180	
Lt.	AAS07180 standard; cDNA; 2037 BP.
XX	AAS07180;
XX	43-UCR-2001 (first entry)
XX	DNA encoding Prosophila beta-carotene dioxygenase (Beta-diox I).
XX	Beta carotene dioxygenase; beta diox I; Prosophila; beta-carotene;
KW	lycopene; beta apocarotenal; beta-ionone; apolycopenal; grain seed, corn,
KM	sui seed; palm, chick-pea; diagenic; therapeutic; ribozyme;
KW	retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;
KM	carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop; ss.
XX	Drosophila melanogaster.
OS	
XX	Key
FT	Location/Qualifiers
FT	CDS
FT	1..1863
FT	/tag= a
FT	/product= "Beta-carotene dioxygenase (beta-diox I)"
PN	M0200148163-A2.
XX	05-JUL-2001.
PD	
XX	27-DEC-2000; 2000MO-EPI3273.
XX	24-DEC-1999; 99EP-0125895.
PR	20-MAR-2000; 2000EP-0105822.
PK	(GPRF-) GFPINOVATION PFLANZENRIECHNOLOGIE GMBH.
PA	
XX	Von Linde J, Vogt K;
EI	WPI: 2001-425657/45.
XX	P-PDB; AA004291.
DR	
PT	Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
XX	(beta diox II) protein that cleaves beta-carotene and lycopene to yield
XX	beta-apocarotenal and beta-ionone, and apolycopenals, respectively -
XX	Disclosure; Fig 6, 116pp; English.
PS	
XX	The sequence represents the coding sequence of Prosophila beta-carotene
CC	dioxygenase (beta-diox I). Beta-diox specifically cleaves beta-carotene
CC	and lycopene to form beta-apocarotenal and beta-ionone, and
CC	apo-lycopenals, respectively. The DNA is useful for transforming grain
CC	seeds, e.g. corn, oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;
CC	edible seeds or seeds with edible parts e.g. chick-peas; potatoes,
CC	carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
CC	for diagnostic and/or therapeutic purposes and for preparing antisense or
CC	ribozyme type therapeutic agents and for detecting any abnormality of
CC	endogenous beta-diox II. The beta-diox II specific antisense
CC	oligonucleotides derived from the DNA sequence are useful for dose
CC	response studies in relevant models of retinoid/vitamin A deficiency
CC	during any stage of an organism's development. The nucleic acids are also

OM protein - protein search, using SW model

Run on: July 15, 2003, 09:04:09 ; Search time 7.2 seconds (without alignments)

973,470 Million cell updates/sec

Title: US-10-053-192-1

Sequence: 1 MEITFNKKKEHPEHKAIV.....MHLIDHMFIPUNDGAETL 520

Scoring table: RUSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 13325620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum OR seq length: 0

Maximum OR seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database:

A-Geneseq_1010022.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	FR	ID	Description
1	28.5	100.0	526	21	AAV97311	Beta, Beta-carotene
2	27.2	96.3	506	21	AAV97314	Beta, Beta-carotene
3	21.07	74.6	547	22	AAV93108	Human protein seqn
4	17.15	60.7	516	22	AAV62810	Amino acid sequence
5	11.42	40.4	549	22	AAV04293	Beta-2, beta-carotene
6	11.16	39.4	579	22	AAV64872	Human PEPAP polypeptide
7	11.12	39.4	532	22	AAV04292	Mouse-2, beta-carotene
8	11.02	39.0	549	22	AAV4454	Human protein seqn
9	11.02	39.0	556	22	AAV04294	Human-2, beta-carotene
10	10.71	37.9	533	22	AAV73955	Canine RPE65, Can

ALIGNMENTS

11	1070	37.9	533	14	AAV44617	Human retinol bind
12	1014.5	35.9	529	21	AAV97315	Beta, Beta-carotene
13	840.5	29.8	523	22	AAV71747	Neoxanthin cleavage
14	840.5	29.8	620	22	AAV94291	Proteolysis beta-carotene
15	840.5	29.8	620	22	AAV62841	Human retinol bind
16	576.5	20.4	357	14	AAV44618	Human retinol bind
17	541	19.2	570	22	AAV12066	Arabidopsis thaliana
18	520	18.4	596	22	AAV12065	Arabidopsis thaliana
19	373.5	13.2	205	22	AAV73956	Canine mutant RPE6
20	311	11.0	538	22	AAV64790	Arabidopsis thaliana
21	310	11.0	538	22	AAV04785	Arabidopsis thaliana
22	307	10.9	501	21	AAV31117	Arabidopsis thaliana
23	307	10.9	538	21	AAV31116	Arabidopsis thaliana
24	307	10.9	539	21	AAV31115	Arabidopsis thaliana
25	276	9.8	605	22	AAV73958	Neoxanthin cleavage
26	276	9.8	605	22	AAV73959	Neoxanthin cleavage
27	263	9.3	577	22	AAV04786	Arabidopsis thaliana
28	263	9.3	577	22	AAV04786	Arabidopsis thaliana
29	260	9.2	599	22	AAV04784	Arabidopsis thaliana
30	260	9.2	599	22	AAV93311	Arabidopsis thaliana
31	247	8.7	297	22	AAV14652	Neoxanthin cleavage
32	239.5	8.5	604	22	AAV04788	Neoxanthin cleavage
33	239.5	8.5	604	22	AAV73959	Neoxanthin cleavage
34	237.5	8.4	589	22	AAV91182	Arabidopsis thaliana
35	233.5	8.2	517	21	AAV31334	Arabidopsis thaliana
36	233.5	8.2	595	21	AAV31333	Arabidopsis thaliana
37	233.5	8.2	595	21	AAV04783	Arabidopsis thaliana
38	233.5	8.2	595	22	AAV73956	Neoxanthin cleavage
39	231.5	8.2	544	22	AAV73953	Neoxanthin cleavage
40	240	7.8	512	22	AAV04787	Neoxanthin cleavage
41	216.5	7.7	404	22	AAV14467	Neoxanthin cleavage
42	207	7.3	583	22	AAV04782	Arabidopsis thaliana
43	197	7.0	446	22	AAV73957	Neoxanthin cleavage
44	176	6.2	431	21	AAV31335	Arabidopsis thaliana
45	133	4.7	110	22	AAV14465	Neoxanthin cleavage

RESULT 1
AAV97311
ID AAV97311 standard; Protein; 526 AA.
XX
AC AAV97311;
XX
CT 07-JAN-2001 (first entry)
XX
DE Beta, Beta-carotene-15,15-dioxygenase.
XX
KW Beta, beta-carotene-15,15-dioxygenase; vitamin A; beta-carotene;
KW transformation; fruit; vegetable; developmental disorder;
KW ophthalmological disorder; antibody; detection; quantification;
KW treatment; therapy.
XX
OS Gallus gallus.
XX
PN EP1031627-A1.
XX
PD 30-AUG-2000.
XX
PP 17-FEB-2000; J00000-010289.
XX
PP 22-FEB-1999; 99EF-0103382.
XX
PA (HOFF) HOPPMANN LA PROTEIN & CO AG F
XX
PI Neumann H, Proctor P, Friedlein AM, Witter CM, Weygen W, Wyss A;
XX
PI Wyss M;
XX
DR WPI: J000-551036/v1
XX
DR N-PSDB: AAA53888.

beta, beta-carotene-15,15'-dioxygenase protein, nucleic acids and antibodies, useful for production of vitamin A from carotene and gene therapy of ophthalmological disorders

Claim 1: Page 12 13: 37pp, English.

Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic conversion of beta-carotene to vitamin A and for raising specific antibodies. Nucleotides encoding all or part of bCOD are useful as primers or probes for specific amplification and/or detection of the gene that encodes bCOD, for isolation of related sequences in other organisms, for determining bCOD levels in humans (to identify subjects requiring vitamin A supplementation) and for detecting mutations in the bCOD gene. The nucleotide coding bCOD may also be used to transform cells, particularly plant cells, to increase their vitamin A contents (especially in fruits and vegetables) and in gene therapy of subjects who have mutated or deleted forms of the bCOD gene and thus low vitamin A levels and susceptibility to developmental or ophthalmological disorders. Antibodies raised against bCOD are used for detection/quantification of bCOD in immunoassays.

Sequence 526 AA:

Query Match 100.0%; Score 2825; DB 21; Length 526,
Best Local Similarity: 100.0%; Prod. No. 2, 2e 247,
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 METTFNNKKEPEPEKAEVGGGLPTWLGVLRLNKGKMHIIIGDKRYNMEGLALLHSF 60
DB 1 METTFNNKKEPEPEKAEVGGGLPTWLGVLRLNKGKMHIIIGDKRYNMEGLALLHSF 60
QY 61 TFKNGEYVYSKYLRSTNYNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 120
DB 61 TFKNGEYVYSKYLRSDYNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 120
QY 121 TDNCLINIMKTGGDYATSENFIRKIDPOTITLADKVDYSKYVAVNATLSHPHDSAGN 180
DB 121 TDNCLINIMKTGGDYATSENFIRKIDPOTITLADKVDYSKYVAVNATLSHPHDSAGN 180
QY 181 HNMGTSTVVKPRTKVTFKIPSSVPEKPKKSKFENHEVWSTSPSLQPSYHSFGI 240
DB 181 HNMGTSTVVKPRTKVTFKIPSSVPEKPKKSKFENHEVWSTSPSLQPSYHSFGI 240
QY 241 TENYIVIEQPEKLDIVKLATAYIRGVNMASSCLSEFKEDKTFHFVDKRTKEVSTKEYT 300
DB 241 TENYIVIEQPEKLDIVKLATAYIRGVNMASSCLSEFKEDKTFHFVDKRTKEVSTKEYT 300
QY 301 DALVIVHHINAYEEDGHVFDIVAYPDNSLYDMPYIKLDRPEVNNKLTSLPTCKRPVY 360
DB 301 DALVIVHHINAYEEDGHVFDIVAYPDNSLYDMPYIKLDRPEVNNKLTSLPTCKRPVY 360
QY 361 PEGYDKDAEVGNSIVKILPTSATAVKREKGSIVYQDELLCEJTELPVAVYDNGKKYKVV 420
DB 361 PEGYDKDAEVGNSIVKILPTSATAVKREKGSIVYQDELLCEJTELPVAVYDNGKKYKVV 420
QY 421 ATEVQWSPVPTKIAKLNVQIKVELHNGEFHFWPSEYIEVPSGLAEELGVLLTVVVSF 480
DB 421 ATEVQWSPVPTKIAKLNVQIKVELHNGEFHFWPSEYIEVPSGLAEELGVLLTVVVSF 480
QY 481 PNRAPELLIDAKTFKELGPATVNMHLDLGMFIPONDGAFIE 526
DB 481 PNRAPELLIDAKTFKELGPATVNMHLDLGMFIPONDGAFIE 526

```

RESULT 2

ID AAV97314 standard; Protein: 506 AA.

AAV97314:

01-JAN-2001 (first entry)

Beta, beta-carotene-15,15'-dioxygenase.
Beta, beta-carotene-15,15'-dioxygenase; vitamin A; beta-carotene; transformation; fruit; vegetable; developmental disorder; ophthalmological disorder; antibody; detection; quantification; treatment; therapy.

Gallus gallus.

EP1031627-A1.

30-AUG-2000.

17-FEB-2009; 2009EP-0102289.

22-FEB-1999; 99EP-0103382.

(HOFF) HOFFMANN LA ROCHE & CO AG F.

bachmann H, Brugger R, Friedlein AM, Wirtz GM, Wozniak W, Wyss A;

WPI: 2000-551036/51.

Beta, beta-carotene-15,15'-dioxygenase protein, nucleic acids and antibodies, useful for production of vitamin A from carotene and gene therapy of ophthalmological disorders

Disclosure: Page 17-18: 37pp, English.

Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic conversion of beta-carotene to vitamin A and for raising specific antibodies. Nucleotides encoding all or part of bCOD are useful as primers or probes for specific amplification and/or detection of the gene that encodes bCOD, for isolation of related sequences in other organisms, for determining bCOD levels in humans (to identify subjects requiring vitamin A supplementation) and for detecting mutations in the bCOD gene. The nucleotide encoding bCOD may also be used to transform cells, particularly plant cells, to increase their vitamin A contents (especially in fruits and vegetables) and in gene therapy of subjects who have mutated or deleted forms of the bCOD gene and thus low vitamin A levels and susceptibility to developmental or ophthalmological disorders. Antibodies raised against bCOD are used for detection/quantification of bCOD in immunoassays.

Sequence 506 AA:

Query Match 95.3%; Score 2723; DB 21; Length 506,
Best Local Similarity: 100.0%; Prod. No. 5, 7e-238;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 10 EEPPEPKAEVGGGLPTWLGVLRLNKGKMHIIIGDKRYNMEGLALLHSF 60
DB 1 EEPPEPKAEVGGGLPTWLGVLRLNKGKMHIIIGDKRYNMEGLALLHSF 60
QY 61 TFKNGEYVYSKYLRSTNYNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 120
DB 61 TFKNGEYVYSKYLRSDYNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 120
QY 121 TDNCLINIMKTGGDYATSENFIRKIDPOTITLADKVDYSKYVAVNATLSHPHDSAGN 180
DB 121 TDNCLINIMKTGGDYATSENFIRKIDPOTITLADKVDYSKYVAVNATLSHPHDSAGN 180
QY 181 HNMGTSTVVKPRTKVTFKIPSSVPEKPKKSKFENHEVWSTSPSLQPSYHSFGI 240
DB 181 HNMGTSTVVKPRTKVTFKIPSSVPEKPKKSKFENHEVWSTSPSLQPSYHSFGI 240
QY 241 TENYIVIEQPEKLDIVKLATAYIRGVNMASSCLSEFKEDKTFHFVDKRTKEVSTKEYT 300
DB 241 TENYIVIEQPEKLDIVKLATAYIRGVNMASSCLSEFKEDKTFHFVDKRTKEVSTKEYT 300
QY 301 DALVIVHHINAYEEDGHVFDIVAYPDNSLYDMPYIKLDRPEVNNKLTSLPTCKRPVY 360
DB 301 DALVIVHHINAYEEDGHVFDIVAYPDNSLYDMPYIKLDRPEVNNKLTSLPTCKRPVY 360
QY 361 PEGYDKDAEVGNSIVKILPTSATAVKREKGSIVYQDELLCEJTELPVAVYDNGKKYKVV 420
DB 361 PEGYDKDAEVGNSIVKILPTSATAVKREKGSIVYQDELLCEJTELPVAVYDNGKKYKVV 420
QY 421 ATEVQWSPVPTKIAKLNVQIKVELHNGEFHFWPSEYIEVPSGLAEELGVLLTVVVSF 480
DB 421 ATEVQWSPVPTKIAKLNVQIKVELHNGEFHFWPSEYIEVPSGLAEELGVLLTVVVSF 480
QY 481 PNRAPELLIDAKTFKELGPATVNMHLDLGMFIPONDGAFIE 526
DB 481 PNRAPELLIDAKTFKELGPATVNMHLDLGMFIPONDGAFIE 526

```



```

DB 301 NAAFTFHHVAVPPIYAVPINSIYPMFTYKLLKLEEVNKKLSIPTCKFVVPPLQYDLAE
OY 370 VGSNLVKLPTSAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYKYYATEVGSPPV 429
DB 361 VGSNLVKLPTSAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYKYYATEVGSPPV 420
OY 430 PTKIAKLNVQTRKEVYHMGEDHQPSEPIYVPSDPAREDEGCVLTCTVASEPKAPFLI 489
DB 421 PTKIAKLNVQTRKEVYHMGEDHQPSEPIYVPSDPAREDEGCVLTCTVASEPKAPFLI 480
OY 490 LDKATFKELGRATVAVEMHDLHGMF 515
DB 481 LDKATFKELGRATVAVEMHDLHGMF 506

RESULT 3
AAB93108
ID AAB93108 standard: Protein: 547 AA
XX
AC AAB93108:
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO: 11972
XX
KM Human: primer: detection; diagnosis; antisense therapy; gene therapy
XX
OS Homo sapiens.
XX
PN EP1074617-A2
XX
PD 07-FEB-2001.
XX
PF 26-JUL-2000; 2000EP-0116125
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0184767.
PR 09-JUN-2000; 2000JP-0247899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ola T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto T;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T.
XX
DR WPI: 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
PS Claim 8; SEQ ID 11972; 2537Pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primer sets are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length

```

```

CC cDNAs easily without any specialised methods. AAH05106 to AAH19028 and
CC AAH13633 to AAH16712 represent human cDNA sequences; AAB92456 to
CC AAB95893 represent human amino acid sequences; and AAH13632 to
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SO Sequence 547 AA:
OY Query Match 74.6%; Score 2107; DB 22; Length 547;
OY Best Local Similarity 70.3%; Pred. No. 3,36-182;
OY Matches 367; Conservative 96; Mismatches 57; Indels 2; Gaps 2;
DB 1 METTNRKKEHPERIKAEVQGLPWLQVLLKNGSGMHLGILKYNHMFQDLALHSL 60
DB 1 MDLFGNPKKLELEVPKAVYNSKIPAWLQGLLPLNPKMHWVESHYNHMFQDLALHSL 60
OY 61 TFRNGEVYRSKYLPSDYNFNTFANPPIVVSFFPTMAVPPCKNIFAKASYLSRTTPE 120
DB 61 TFRNGEVYRSKYLPSDYNFNTFANPPIVVSFFPTMAVPPCKNIFAKASYLSRTTPE 120
OY 121 TDNCLINIMKTGGDYATSETNIRKIDPOTLEFLDKVDYSKYVAVNLAATSHPHYSAGN 180
DB 121 TDNCLINIMKTGGDYATSETNIRKIDPOTLEFLDKVDYSKYVAVNLAATSHPHYSAGN 180
OY 123 TENCLINIMKTGGDYATSETNIRKIDPOTLEFLDKVDYSKYVAVNLAATSHPHYSAGN 180
DB 123 TENCLINIMKTGGDYATSETNIRKIDPOTLEFLDKVDYSKYVAVNLAATSHPHYSAGN 180
OY 124 LNMGSIVKGGPKRYVLPKIDSSVPEKEKKSCGFHLVGSISPSRIIDHSYHSFC 239
DB 124 LNMGSIVKGGPKRYVLPKIDSSVPEKEKKSCGFHLVGSISPSRIIDHSYHSFC 240
OY 240 ITENYIVLPPFKLIDYKLAIVYGVNMAQSLSPFKELKIMHEHVEFTKREVSSTREY 299
DB 240 ITENYIVLPPFKLIDYKLAIVYGVNMAQSLSPFKELKIMHEHVEFTKREVSSTREY 299
OY 241 VFNVNVITLPDPPLDILKMATVIRNSMASCILAPPERKTYHITDQEPDQVCTKY 300
DB 241 VFNVNVITLPDPPLDILKMATVIRNSMASCILAPPERKTYHITDQEPDQVCTKY 300
OY 300 TDALVLYHINAYEEDGHVVDIVAYKUNSLYDVFYLLKIDKDEVNKKITSIPTCKREV 359
DB 300 TDALVLYHINAYEEDGHVVDIVAYKUNSLYDVFYLLKIDKDEVNKKITSIPTCKREV 359
OY 360 VPLQYDKAEVGSNLVAVL-PTSAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYK 418
DB 360 VPLQYDKAEVGSNLVAVL-PTSAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYK 418
OY 361 VFLHVDKAEVGSNLVAVL-PTSAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYK 420
DB 361 VFLHVDKAEVGSNLVAVL-PTSAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYK 420
OY 470 VYATFVNSVFTKTAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYK 478
DB 470 VYATFVNSVFTKTAIYAVKEKDSIYCOPELLCGEIE PPNVYFVNSKKYK 478
OY 479 SPNNKAPFLIITDATEREYIGPATVAVEMHDLHGMF 520
DB 479 SPNNKAPFLIITDATEREYIGPATVAVEMHDLHGMF 520
OY 481 TDPQKLPPLIITDATEREYIGPATVAVEMHDLHGMF 522
DB 481 TDPQKLPPLIITDATEREYIGPATVAVEMHDLHGMF 522

RESULT 4
AAB6840
ID AAB6840 standard: Protein: 516 AA.
XX
AC AAB6840:
XX
DT 17-SEP-2001 (first entry)
XX
DE Amino acid sequence of a beta-carotene dioxygenase (beta-dio).
XX
KW Beta carotene dioxygenase; beta-dio; beta carotene; vitamin A aldehyde;
XX transgenic plant.
XX
US Danto ferro.
XX
IN W0200146162-A2.
XX
PD 05-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-EP13144.
XX
PR 24-DEC-1999; 99JP-0135895.
PR 20-MAY-2000; 2000JP-0135822.
XX

```

PA (GREE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.
 XX
 PI Von Lintig J, Vogt K;
 XX WPI: 2001-441713/47.
 DR N-PSDB: AAH42170.
 XX
 PT Dioxigenase (beta-diox) protein that cleaves beta carotene to form
 PT vitamin A aldehyde, and polynucleotides encoding them useful for
 PT producing transgenic bacteria, fungi, plants expressing the polypeptide
 PT
 PS Claim 7; Page 86-87; 87pp; English.
 XX
 CC The present sequence represents a beta-carotene dioxigenase (beta-diox);
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic
 CC plants. The transgenic plants have improved nutritional quality or
 CC physiological condition and accumulate vitamin A aldehyde and can take
 CC up beta-carotene from the medium. Expression systems encoding beta-diox
 CC are useful in the study of beta-diox activity. Identification of cDNAs
 CC encoding beta-diox allows the physiological characterization of
 CC mammalian vitamin A metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.
 XX
 SQ Sequence 516 AA;
 Query Match 60.7%; Score 1715, DB 22, Length 516;
 Best Local Similarity 59.7%; Pred. No. 1,1e-146;
 Matches 310; Conservative 87; Mismatches 114; Indels 8; Gaps 5;
 OY 1 METFNKKKEHPPIKAEVQGLPTMLQGVLLRNGPCHMTTIDTKYNNHMFQGLALHSF 60
 DB 1 MVDYGNKKKEHPPIETEVKQSIPEVQGTILKNGPMHVSDELTYNHMFQMLALHSF 60
 OY 61 TFKNGEYYSKYTSRSTYCNTEANRIVYSEGTATYDPCCKNTFAKFSLSHTPEF 120
 DB 61 AINKEGYTYSRLRGDTTNSMOMANIVYSEGTATYDPCCKNTFAKFSLSHTPEF 120
 OY 121 TPNCLINIMKTGDYVATSEINPINKIDPYLETLKLVYSKYAVANALISHGHVSACN 180
 DB 121 TDNCGNNIIRKYNDGFHATSETNIRKIDPYLTETOKRIDYLYKLPYSIVASHTHDKEN 180
 OY 181 ILNMGTSIVDKGRKRYVLEKIP--SSYPEKSKCKKHLFVSCISPSKSLDPSYHSFG 239
 DB 181 SYSNGTIAKGGTKTKMLKRVPPSPSPSP--LKSAAVCTLPSPSLTPSYHSFG 237
 OY 240 ITENYVIEIPEPKLDIVKLATAYIRGVNMAASCLSEHKEDKTFHFVDRKTKKFSSTFY 299
 DB 238 MTUNYFIFIQPLKLDILKMATVILFVNSMASCMEKHPEDSLTILHLENTKKEVATKTY 297
 OY 300 TDAIVLVIHNINAEEDGHVVDIVATKDSNIDMFYLLKLDKDFEVNKLISIPCKKRV 359
 DB 298 TDAMTYVHOVNAFEDKHVVDIVAYDNNLVEFFELNFKFTMATN-LYCKPKFTFV 356
 OY 360 VPILOYKDAEVSNLKLP-TSATVAKKDGSIYCOOPLIIGGIELPRVNYNKKKKY 418
 DB 357 FPLI--SIQGTGEDGLKLYTTASAKKEDKIMCGEVLGCEVELPRINTFNCKKTY 414
 OY 419 VATEVWVSFPTKIAKLNVQREVLHMGEDHQPSELPVSPDAREDECVLTGVV 478
 DB 415 SYMCVDSVPATRIKFPADFKQQLEMKGDGKPSAFEPFIPPCAVDEDDCVLTVIN 474
 OY 479 SEFNKAPHLILDAKTFKELGRATVNEHMLDGMETP 517
 DB 475 NKPIQGGFLVLDAKSFKEIARACLDELHMDHGFIP 513

AC AA04293;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Zebra-2, beta-carotene dioxigenase (Beta-diox II).
 XX
 KM Beta-carotene dioxigenase; beta-diox II; zebra-2; beta-carotene;
 KM lycopene; beta-apocarotenal; beta-ionone; apolyoponals; grain seed; corn;
 KM 511 seed; palm; chick pea; diapaestic; therapeutic; lycopene; zebrafish;
 KM retinoid/vitamin A deficiency; beta-diox II; transgenic nutrition;
 KM carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop.
 XX
 OS Brachydanio rerio.
 XX
 IN WU00148163-A2.
 XX
 PD 05-JUL-2001.
 XX
 XX 27 DEC 2000, 2000W, EF13-73.
 XX
 XX 24-DEC-1999, 2000P-0124805.
 XX 20 MAR 2000, 2000P-0124822.
 XX
 PA (GREE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.
 XX
 PI Von Lintig J, Vogt K;
 XX WPI: 2001-425657/45.
 DR N-PSDB: AAS07195.
 XX
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxigenase
 PT (beta-diox II) protein that cleaves beta carotene and lycopene to yield
 PT beta-apocarotenal and beta-ionone, and apolyoponals, respectively.
 PS Claim 2; Fig 14; 116pp; English.
 CC The sequence represents the amino acid sequence of zebra 2, beta-carotene
 CC dioxigenase (beta-diox II). Beta-diox specifically cleaves beta-carotene
 CC and lycopene to form beta-apocarotenal and beta-ionone, and
 CC apolyoponals, respectively. The DNA is useful for transforming grain
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or
 CC ribozyme type therapeutic agents and for detecting any detectability of
 CC endogenous beta-diox II. The beta-diox II specific antisense
 CC oligonucleotides derived from the DNA sequence are useful for dose
 CC response studies in relevant models of retinoid/vitamin A deficiency
 CC during any stage of an organism's development. The nucleic acids are also
 CC useful as probes and as a guideline to define new PCR (polymerase chain
 CC reaction) primers for the cloning of substantially homologous DNA
 CC sequences from other sources. The nucleic acids are also useful for
 CC determining the presence or quantity of beta-diox II nucleic acid and
 CC useful for increasing or decreasing the amount of beta-diox II levels in
 CC a cell or tissue which can modify the level of vitamin A and other
 CC retinoids. Antibodies are useful for studying beta diox II localisation,
 CC screening of an expression library to identify nucleic acids encoding
 CC beta-diox II or the structure of functional domains. The transgenic
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
 CC organisms have improved nutritional quality or physiological condition
 CC and accumulate improved metabolites of carotenoid/retinoid pathways such
 CC as vitamin A aldehyde and retinoic acid. beta-carotene or take up beta
 CC carotene from the medium. Expression systems encoding beta-diox II are
 CC useful in the study of beta-diox II activity. Identification of cDNAs
 CC encoding beta-diox II and II allows the physiological characterization of
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.
 XX
 SQ Sequence 549 AA;
 Query Match 40.4%; Score 1142.5, DB 22, Length 549;

AA04292 standard; Protein; 532 AA.
 AA04292;
 23-OCT-2001 (first entry)
 Mouse-2, beta-carotene dioxygenase (Beta-diox II).
 beta-carotene dioxygenase; beta-diox II; mouse-2; beta-carotene,
 lycopene, beta-apocarothenal; beta-ionone, apolycopene; grain seed; corn;
 oil seed; palm; chick-pea; diagnostic; therapeutic; ribozyme;
 retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;
 carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop.
 Mus musculus.
 M0200148163-A2.
 05-JUL-2001.
 27-DEC-2000; 2000MO-EP13273.
 24-DEC-1999; 99EP-0125895.
 20-MAR-2000; 2000EP-0105822.
 (GREE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH
 Von Lintig J, Vogt K;
 WPI: 2001-425657/45.
 N-PSDR: AAS07194.
 Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
 (beta-diox II) protein that cleaves beta-carotene and lycopene to
 beta-apocarothenal and beta-ionone, and apolycopene, respectively
 Claim 2; Fig 14; 116pp; English.
 The sequence represents the amino acid sequence of mouse-2, beta-carotene
 dioxygenase (beta-diox II). Beta-diox specifically cleaves beta-carotene
 and lycopene to form beta-apocarothenal and beta-ionone, and
 apolycopene, respectively. The DNA is useful for transforming grain
 seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;
 edible seeds or seeds with edible parts e.g. chick-peas; potatoes,
 carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
 for diagnostic and/or therapeutic purposes and for preparing antisense or
 ribozyme type therapeutic agents and for detecting any abnormality of
 endogenous beta-diox II. The beta-diox II specific antisense
 oligonucleotides derived from the cDNA sequence are useful for drug
 response studies in relevant models of retinoid/vitamin A deficiency
 during any stage of an organism's development. The nucleic acids are also
 useful as probes and as a guideline to define new PCR (polymerase chain
 reaction) primers for the cloning of substantially homologous DNA
 sequences from other sources. The nucleic acids are also useful for
 determining the presence or quantity of beta-diox II nucleic acid and
 determining presence and amount of beta-diox II. The polypeptide is also
 useful for increasing or decreasing the amount of beta-diox II levels in
 a cell or tissue which can modify the level of vitamin A and other
 retinoids. Antibodies are useful for studying beta-diox II localisation,
 screening of an expression library to identify nucleic acids encoding
 beta-diox II or the structure of functional domains. The transgenic
 fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
 organisms have improved nutritional quality or physiological condition
 and accumulate important metabolites of carotene/retinoid pathways such
 as vitamin A aldehyde and retinoic acid. beta-diox II can take up beta-
 carotene from the medium. Expression systems encoding beta-diox II are
 useful in the study of beta-diox II activity. Identification of cDNAs
 encoding beta-diox I and II allows the physiological characterisation of
 mammalian carotene/retinoid metabolism. Vitamin A production in crops and
 microorganisms can be achieved by transforming the organisms or crops
 with the cDNAs.
 Sequence 532 AA;

Query Match 39.48; Score 112.5; DN 22; Length 542;
 Best Local Similarity 43.58; Pred. No. 5.4e-92;
 Matches 228; Conservative 94; Mismatches 100; Gaps 10;
 QY EEHPEPIKAEVGGOLPTLGLGVLKNGPGMHTIGDKRYNHNHPLALHSFTKKNLEVVY 69
 10 EETLSASAVPBGHIEPLMLNYLLPVSPGFEEFGKDRYNNHPLMGALHDPGMEKGVY 78
 DB 19 EETLSASAVPBGHIEPLMLNYLLPVSPGFEEFGKDRYNNHPLMGALHDPGMEKGVY 78
 VY 70 KSKTLKSLIYNCLNLEAKRYVSLDGLMAYIDFKNIIAAGSYLSITFTHTTRLLHII 129
 DB 79 KSKFLGNTYKANSAGSPVIVSPGGLAIPIPKSTFFPFMSFPF--PTMTDINNVNV 146
 QY 130 KTGDYVATSEFTNFKIDPQTELEKVSRYVAVNATSPRHSAGNINMISIV 189
 DB 137 QKGDYVATSEFTNFMKNVDIEMLFRTEKVMSEFLAVNATAPHTDPAKIVNMGNSYV 196
 QY 190 DKGRKRYLKRIPSSVPEKSKSCPKHLEWGSIPSRSLDPSYHSFGITENVYVIE 249
 DB 197 PRG-SCYNIIRVP--PKRKEPETIHGAQVLCSIASTEKKMPSYHSFGMTKNYIIVE 252
 QY 250 QPFLDIYKLAIAVIRKVNMSASLSFHKKEKTFNHFVPIKRVSKRYTIALVYHNI 409
 DB 253 QPVMFLMFLIISKIPKPAIGLWELGVYRHHVWLFPHGLGAGVYSHMILLYHQ 412
 QY 310 NAVEEPCHVVEDYVAVPNSLYDMYLRK-----LPTFTYNNPLISFTYRFTVYV 462
 DB 313 NAFRGGCIVIDICCGHRSKSLDYGNLWAGAGHGVYVET--KANSFP RRPVLDL 468
 QY 363 QYLDAAVGSMLKLP LSAIAVRLKNSLYGHLDT GILLRVRD YNG 414
 DB 369 DVSVAALGKNNLSPLSYSSASAVKGGHGLMGLFENHMDLEEDGIEFPGLNGLPNG 428
 QY 414 KKYVYVATVQWSPVTRKAKINVCREVIHMGEPHWPSEPTFVSPMAFEDGVV 474
 DB 429 KKYFFYGGCPR-HLVDSLLIKVDYTNKTLRVRKEGCPFPSPGVYVYVAGALRHSVYL 487
 QY 474 TCVVVSEPNKAPFLLLIDAKTFRELSPATVNMHMLDLNGKIP 517
 DB 488 SVVITPQSRNELLVLAASFELGRAEVPGVMPGPGHTRVP 531
 RESULT 8
 AAB93458
 ID AAB93458 standard; Protein; 539 AA.
 AC AAB93458;
 CT 24-JUN-2001 (first entry)
 Human protein sequence SEQ ID NO:12718.
 Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 Homo sapiens.
 EP1074617-A2.
 PD 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-000253.
 XX 13-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241809.
 PA (HELIX) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Onuki T;
 XX

DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 PS Claim 8: SEQ ID 12718; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH0156 to AAH1368 and
 CC AAH1363 to AAH18742 represent human cDNA sequences; AAH9249 to
 CC AAH95893 represent human amino acid sequences; and AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 539 AA:

Query Match 39 0%: Score 1102.5; DH 22; Length 539;
 Best Local Similarity 44.3%; Pred. No. 4,4e-91;

Matches 232; Conservative 85; Mismatches 118; Indels 29; Gaps 10;

DB 10 EEHPPIKAEVGGOLPTLGGVLLBNPGMHHTIGDTKYNNHFDGLALHSTFKNGEYVY 69
 DB 28 EEAHPGQISARWCHPEKMLNSLKLKCHKEPKDKMHWEDMALHOFKMAKGVITY 87
 QY 70 FSKYPTPTPTNTANFTVSEPTMAVPECEKIFAKAKS YLHIIIEETELIN 127
 DB 88 RSKPLQSTQYKANSKKNIVISKRTALPPPKNVEFPKSPRELPGKAAAMTNTNN 147
 QY 128 IMKTDDYYATSETNFKRIDQTLTLDKDYKSYAVANLASHPHYDSAGNLLNGSTS 187
 DB 148 YVRVKGDDYVLCETEFNMKNKVDLETLEKEDKMSKFTAVN-CATNHPHYDPGTAYMNGNS 207
 QY 188 IVDPKPTVYVLFKTISSVPEKEKKKSPKHI EWGCSIPSEGLDPSYHSGTIENTIVE 247
 DB 208 FEPYVC-FSKYVLRV--FEFVCLGETIHVAVGCTIASTFKSPSYHSRGMILNIF 269
 QY 248 IEOPKRLIVKLATVYIRGVNMASTLSFHEEKTWIEFVLEKIKKEVSIKMYIALALYH 307
 DB 264 ITPQLKMNIMKATISKTKGKAFSGTISMEVQCNIRFHVVEKPTQILLPSYNSPVYTH 323
 QY 308 HINAVEEDGCHVHIVAVRNSIVMPYTKIKLK--FEVNNKLITSTPKFVAVI 362
 DB 324 QINAEEDGCVILIDCCZCNGFTLEVYCIQNLFRAGETLIDVHNSAKSPF--FEFVLP 381
 QY 363 QYDKAEFGVSNIVKLP-TSATAVNKKKGSIVQGEPLIKR-----RIELPKVAVYD-YNG 414
 DB 382 NWSLNAFECODNLSPLSTYSASAVKQADCIWCHSHENLMDULEKEGGIEFQIYTHFSG 441
 QY 414 KRYKYVYATEVQMSVPPTKIATINQTEVLHMGEDPHWSEPTFVSPFAFEESVVL 473
 DB 442 KRYHFFYGGGPR-HI VGSLSLIKVVV-----WEPDGYPRSPFVAPATNFRPDGVIL 494
 QY 474 TCVVVSSEKKAFFILLDAKTFKELGSAIVNENHLLHGMFIP 517
 DB 495 SVVITPNDNSNELLVLDAKNFELIGRAFEVVDGMPYGGHGTIFIP 538

RESULT 9
 AAU04294
 ID AAU04294 standard; Protein; 556 AA.
 XX
 AC AAU04294;
 XX
 DI 23-CC1-2001 (first entry)
 XX
 DE Human-2, beta-carotene dioxygenase (beta-diox II).
 XX
 KW Beta-carotene dioxygenase; beta-diox II; human-2; beta-carotene;
 KW lycopene; beta-apocarotenal; beta ionone; apolycopenal; grain seed; corn;
 KW oil seed; palm; chick-pea; diagnostic; therapeutic; ribozyme;
 KW retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;
 KW carotene; retinoid pathway; vitamin A aldehyde; retinoic acid; crop.
 XX
 US Home sapiens.
 XX
 IN WC200142163-A2.
 XX
 PV 05-JUL-2001.
 PD 27-DEC-2000; 2000M3 EP13273.
 XX
 XX 24-DEC-1999; 1999-01-5895.
 PV 20-MAR-2000; 2000EP-0105822.
 XX
 XX (GFRF-) GFRFN-VATION PLANTENKLEINIGUNG GMBH.
 PA
 XX Von Lintig J, Vogt K;
 PI
 XX WPI: 2001-425657/45.
 DR
 XX N PSDB: AAS07146.
 XX
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
 PT (beta-diox II) protein that cleaves beta carotene and lycopene to yield
 PT beta-apocarotenal and beta-ionone, and apolycopenal, respectively.
 XX
 XX Claim 4, Fig 14, 11ep; English.
 CC The sequence represents the amino acid sequence of human 2, beta-carotene
 CC dioxygenase (beta diox II). Beta diox specifically cleaves beta-carotene
 CC and lycopene to form beta-apocarotenal and beta-ionone, and
 CC apolycopenal, respectively. The DNA is useful for transforming grain
 CC seeds, e.g. corn; oil seeds, e.g. brassica seeds; edible seeds e.g. palm;
 CC edible seeds or seeds with edible parts e.g. chick-peas, potatoes,
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or
 CC ribozyme type therapeutic agents and for detecting any abnormality of
 CC endogenous beta-diox II. The beta-diox II specific antisense
 CC oligonucleotides derived from the DNA sequence are useful for dose
 CC response studies in relevant models of retinoid/vitamin A deficiency
 CC during any stage of an organism's development. The nucleic acids are also
 CC useful as probes and as a guideline to define new PCR polymerase chain
 CC reaction primers for the cloning of substantially homologous DNA
 CC sequences from other sources. The nucleic acids are also useful for
 CC determining the presence or quantity of beta diox II nucleic acid and
 CC determining presence and amount of beta-diox II. The polypeptide is also
 CC useful for increasing or decreasing the amount of beta diox II levels in
 CC a cell or tissue which can modify the level of vitamin A and other
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,
 CC screening of an expression library to identify nucleic acids encoding
 CC beta-diox II or the structure of functional domains. The transgenic
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
 CC organisms have improved nutritional quality or physiological condition
 CC and accumulate important metabolites of carotenoid/retinoid pathways such
 CC as vitamin A aldehyde and retinoic acid. beta-carotene or take up beta-
 CC carotene from the medium. Expression systems encoding beta-diox II are
 CC useful in the study of beta-diox II activity. Identification of cDNAs
 CC encoding beta-diox I and II allow the physiological characterisation of
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and

CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.

XX Sequence 556 AA:

Query Match 39 08 Score 1162, Pos 22, Length 556;
 Host Local Similarity 44.18; Pred. No. 4.6e-91;
 Matches 231; Conservative 87; Mismatches 183; Indels 23; Gaps 9;

```

UY 10 EEHEPPIKAEEVQGLPTWLGVLKNGCMHTIGTRKYNHWEGLALLHSFTKKEEYVY 69
DB 39 EEARGISARVWGHPFWMNGSLRLIQPKFERGKKNYMFQMLLQFPMKACTVTV 98
OY 70 RSKYLSRDTYNCNIEANPIVVSFEGTMAVPPCKNIFAKAFS--YLSHTIPEPTDCLIN 127
DB 99 RSKFLQSDTYKANKAKRIYISSEGTALPDPCKNFERMSFELPGRKAAANTDDTNNV 158
OY 128 IMKTGGDYATSEINFRKIDPOTLETLKVDYSKVAVANLATSHPHYDSAGNLMGTS 187
DB 159 YVRKKGYYLCTETENFNKKVDIETLEKTEKVDMSKFLAVGATAHHPYDPDGTAYNMGNS 218
OY 188 IVDGPTKRYVLFKIPSSVPEKEKKSCFKLEEVGSIPESSLQSYHSFGITENYIVF 247
DB 219 FGPRG-FSYKVIKRP---PREVDGETIHGVQVICSIASTEKCKPSYHSFGITRYITF 274
OY 248 IEGPFLDIKLAATAYIRGVNMAASCLSFHKEDKTMHFVDKTKKEVSTKRYTDAVLYH 307
DB 275 IEGPLKNNIMKIAISKIRGKAFSGISMEPQCTPFHVEKRTGQLPGRYYSKPFVTH 334
OY 308 HINAEKHGHVVDIVAYRNSLSLYDMFYLLKLLK-----DFEYNNKLTSIPCKRNVVL 362
DB 435 QINAEFGQVYIIMFGQDMGKPIFYVQIQNLPRKAGPGLNQVHNSAKSP--RPFVLP 392
OY 463 QYDKDAEVSNLVLP--TSATAVKEKDSIYCOPELLCE-----GIELPRVND-YNG 413
DB 493 NVSLNADEGNLSPLSTYSASAVKQACGTICSHENLHODELEKEGIEFPQIYYDRFSG 452
OY 414 KRYRYVATEVQMSPVTKIAKLNVOTKEVLAHMGEDHCWSPRTIVPSPDAREDEGVVL 473
DB 453 KRYHFEYGGCPR-HLVGDSLIKVDYNNKTLKVRDEGFYSPSEVFPARGTNEEDGVIL 511
OY 474 TGVVSEPNKAPFLILDAKTFKEIGATVNVEMHLDLHGMPF 517
DB 512 SVVITTRQNSNELLVLDKAKNFEIGRAFPVQMPYGFHGTIFP 555

```

RESULT 10

AAB73955

1D AAB73955 standard; Protein: 533 AA.

XX AAB73955;

DT 29-MAY-2001 (first entry)

XX Canine RPE65.

XX Canine: dog; RPE65, canine retinal pigment epithelium 65; RPE65;

XX congenital stationary night blindness; CSNB; mutation detection.

XX Canis familiaris.

XX US6201114-H1.

XX 13-MAR-2001.

XX 30-AUG-1999; 99US-0385259.

XX 06-OCT-1998; 98US-0103219.

XX (CORNELL) CORNELL RES FOUNDD INC.

XX Aquifre GD, Acland GM, Ray K;

DB WPI: 2001-265162/27.

DB N-PSDB: AAB79560.

XX Novel nucleic acid molecule encoding canine retinal pigment epithelium
 PT 65, where presence of mutation in one or both alleles is indicative of
 P1 a carrier of, or dog affected with, congenital stationary night
 XX blindness.

PS Claim 2: Fig 1A-1B; 25pp; English.

XX The present sequence is canine retinal pigment epithelium 65 (RPE65). The
 CC gene encoding this protein is useful for identifying dogs which are
 CC genetically normal, or are carriers of, or affected with congenital
 CC stationary night blindness (CSNB). Carriers for CSNB are identified
 CC through the detection of an AACA deletion at nucleotides 487-490 in the
 CC canine RPE65 gene. This allows a breeder to eliminate the carrier from
 CC the breeding stock or to breed carriers with genetically normal dogs.

XX Sequence 533 AA:

Query Match 37.98; Score 1071; DB 22; Length 533;
 Host Local Similarity 42.28; Pred. No. 3.1e-88;
 Matches 222; Conservative 92; Mismatches 188; Indels 24; Gaps 12;

```

OY 2 ETIFRNKKEPPIKAEEVQGLPTWLGVLKNGCMHTIGTRKYNHWEGLALLHSFT 61
DB 17 EIV-----ELSSPLAHVIGRPLMLSSLLSCGDFEVSSEDEYHFEQALLHRED 71
OY 62 FKRGEEVYRKKYLLPSDTYNCNIEANPIVVSFEGTMAVPPCKNIFAKAFS--YLSHT 121
DB 72 FKRGEEVYRKKYLLPSDTYNCNIEANPIVVSFEGTMAVPPCKNIFAKAFS--YLSHT 129
OY 122 DNCLINIMKTGGDYATSEINFRKIDPOTLETLKVDYSKVAVANLATSHPHYDSAGN 181
DB 130 DNALVNVYRPGCPVYAACTETNPITKINPTIETKGVQVQNVSVNGATAHHPYDPDGT 189
OY 182 LNMGTSIVDGKRIKRYVLFKIPSSVPEKEKKSCFKLEEVN--LSKSLIGSYHSFGIT 241
DB 190 YINQNCFGKNESIAVNVKIF--FLQADKEDDISKSEVVVQFSDPREKPSYVHSPDT 246
OY 242 ENYIVFEQPKLDIVLAZAY--LPGVNMASVCLSFHKEDKTMHFVDKTKKEVSTKRYT 400
DB 247 PNIVVPEVTVKINILKFLSSWSLWQANIMDETFSSNLMVWHLIAHAKRKNVIMNNYRT 406
OY 301 DALVLYHINAEYEDGHVEFDIVAYRP--NSLYDMFYLLKLLKDFE--VNNKLTSIPCK 456
DB 307 SSFNLFFHHTIYETNEFLVLDLCMKRGFEVYVYVLAHNPENMFEVKKNAKAPQPEVR 466
OY 357 PFVVPGLQYDKDAEVSNLVLP--TSATAVKEKDSIYCOPELLCE---JELPRVND- 410
DB 367 RSVLPLENDK-AOTGKMLVTLPTNTATVATLRSDETTWLEFEVLSFGKQAFPEPQINQK 425
OY 411 YNSKRYRYVVAIEVQMSPVTKIAKLNVOTKEVLAHMGEDHCWSPRTIVPSPDAREDE 470
DB 426 YGKRPTYTAVAGGLNHF--VPDRICKLVNKTRETVMWQGEDPSVSEPTFVSHDLEEDG 484
OY 471 VVLTCVNVSEPNKAP--FLILDAKTFKEIGATVNVEMHLDLHGMPF 515
DB 485 VVLSVYVSPGAGQKPAVLLLNKDLSEVAAAEVETINIPVTRHGLE 530

```

RESULT 11

AAR44617

1D AAR44617 standard; Protein: 533 AA.

XX AAR44617;

DT 02 JUN 1994 (first entry)

XX Human retinol binding protein receptor.

XX Retinol binding protein receptor; retinitis pigmentosa.

OS Homo sapiens.
 XX MO9323538-A.
 XX 25-NOV-1993.
 XX
 XX 14-MAY-1993: 93WO-US04586.
 XX
 XX 15-MAY-1992: 92US-0883539.
 XX
 XX (LUDWIG) LUDWIG INST CANCER RES.
 XX
 XX Bavi CO, Eriksson U, Simon A;
 XX WPI: 1993-186570/48
 XX N-PSDB: AA053070.
 XX
 XX New retinol binding protein receptor and homologue coding nucleic
 XX acid molecule - useful for diagnosis and treatment of retinoid
 XX linked pathological conditions, for hybridization in stringant
 XX conditions and treating retinitis
 XX
 XX Claim 15, Page 21-24; 44pp; English.
 XX
 XX The protein (SFO, ID NO. 1) may be used to diagnose and treat
 XX retinoid linked conditions such as inability to accumulate
 XX retinoids in the eye, leading to retinitis pigmentosa.
 XX
 XX Sequence 533 AA;

Query Match 37.9%; Score 1070; DB 14; Length 533;
 Best Local Similarity 42.0%; Pred No. 3.9e-88;
 Matches 221; Conservative 92; Mismatches 189; Indels 24; Gaps 12;

2 ETITPNKFEPIKAEVQGLPTW:CGVLLFNGFSGMHTGCTKRYNMFQDLALNSFI 51
 17 EIV-----EELSSPLIAHTGRIPMLTGLRCQPGI FEVQSEFFYHIFQGLIHKFD 71
 62 FKGEVYYSKSKYLRSTYKNCLEANRIYSEGTMAVPDCKNIFAKAFSLSHITPEFT 121
 72 FREGHTVIRPFTIDAVFAMTEKPIVITTEGTAFFEDPKNIFSPFFSFSV--EVT 129
 122 DNLCLINMKGGDYVATSENFIRKIDVLTLETLAKVZYSKVAVNAIASHIHVSANI 181
 130 DNALVNIYVGEYVYACTETNITVKNPETLETINQVLCNVSVNGATAPHIENDCTV 189
 182 LMGSTIVDKGRTKYVLRKISSVEKEKKSKTRHLEVVSIPSSSLQPSYHSFST 241
 190 YNIGNCFGNFSLAIVNIVKIP--PLOADKEDPISKSEIVVOFPCSDRFKPSYHSFST 246
 242 ENYIVFIEDPFIKIDYKLTAY-IRGVNNASCLSFHKEDKTMFHFVDRKTRKSTKFT 300
 247 PNIVIVETPVKINLFKFLSSMSLMGANYMCFESNELTMGVALIADKKRKYINNKRT 306
 301 DALVIVHINAYEEDGHVVDIAVARD--NSLYDMFYKLLDKDF--VNNKLISITLCK 356
 307 SFENLFRHINTVEEHEFLVDLQCKJFFVYNYLYANLRFNMHFFVKKNAPKAPQEVF 366
 357 KRVNPLINIK-ADTKNLTATTPNTATATLSTHTIMEFPVIFSGPQAFEPFQINQK 425
 411 YNGKRYVYATEVQSPVETRIAKLNQSTKEVILMGSTHWPSEFIVPSEFLAEEDES 476
 426 YGKRKYTVAYIGLNIHF-VPDKLKIKINATKRETWMQEDRSYSPRTIVSHPTALEDR 484
 471 VITVIVVSFPKAP-FITIDAKTFKFGATVAVENHIGI HSNF 515
 485 VLVSVVSPGAGQKPAYILILNAKDLSEVAPAFVEINIPVTFHDLF 530

RESULT 12
 AA97315

ID AAY97315 standard; Protein; 529 AA.
 XX
 XX AAY97315;
 AC
 XX 03-JAN-2001 (first entry)
 DT
 XX
 XX Beta, beta-carotene-15,15-dioxygenase.
 DE
 XX
 XX Beta, beta-carotene 15,15-dioxygenase; vitamin A; beta-carotene;
 XX transformation; fruit; vegetable; developmental disorder;
 XX ophthalmological disorder; antibody; derivation; quantification;
 XX treatment; therapy.
 XX
 XX Bos taurus.
 XX
 XX EP1031627-A1.
 XX
 XX 30-AUG-2000.
 XX
 XX 17 FEB-2000; 2000EP 0102289.
 XX
 XX 22-FEB-1999; 99EP-0103382.
 XX
 XX (HOF) HOFMANN LA ROCHE & CO AG F.
 XX
 XX Bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;
 XX WISS M;
 XX WPI: 2000-551036/51.
 XX
 XX Beta, beta-carotene 15,15'-dioxygenase protein, nucleic acids and
 XX antibodies, useful for production of vitamin A from carotene and gene
 XX therapy of ophthalmological disorders
 XX
 XX Disclosure, Page 19-20, 37pp; English.
 XX
 XX Beta, beta-carotene-15,15'-dioxygenase (bcoP) is used for enzymatic
 XX conversion of beta-carotene to vitamin A and for raising specific
 XX antibodies. Nucleotides encoding all or part of bcoP are useful as
 XX primers or probes for specific amplification and/or detection of
 XX the gene that encodes bcoP, for isolation of related sequences in
 XX other organisms, for determining bcoP levels in humans (to identify
 XX subjects requiring vitamin A supplementation) and for detecting
 XX antibodies in the bcoP gene. The nucleotide encoding bcoP may also
 XX be used to transform cells, particularly plant cells, to increase
 XX their vitamin A contents (especially in fruits and vegetables) and
 XX in gene therapy of subjects who have mutated or deleted forms of the
 XX bcoP gene and thus low vitamin A levels and susceptibility to
 XX developmental or ophthalmological disorders. Antibodies raised
 XX against bcoP are used for detection/quantification of bcoP in
 XX immunoassays.
 XX
 XX Sequence 529 AA;

Query Match 35.9%; Score 1014.5; DB 21; Length 529;
 Best Local Similarity 42.3%; Pred No. 4.1e-83;
 Matches 225; Conservative 91; Mismatches 184; Indels 53; Gaps 20;

10 SEHPEPIKAEVQGLPTW:CGVLLFNGFSGMHTGCTKRYNMFQDLALNSFI 51
 17 EIV-----EELSSPLIAHTGRIPMLTGLRCQPGI FEVQSEFFYHIFQGLIHKFD 71
 62 FKGEVYYSKSKYLRSTYKNCLEANRIYSEGTMAVPDCKNIFAKAFSLSHITPEFT 121
 72 FREGHTVIRPFTIDAVFAMTEKPIVITTEGTAFFEDPKNIFSPFFSFSV--EVT 129
 122 DNLCLINMKGGDYVATSENFIRKIDVLTLETLAKVZYSKVAVNAIASHIHVSANI 181
 130 DNALVNIYVGEYVYACTETNITVKNPETLETINQVLCNVSVNGATAPHIENDCTV 189
 182 LMGSTIVDKGRTKYVLRKISSVEKEKKSKTRHLEVVSIPSSSLQPSYHSFST 241
 190 YNIGNCFGNFSLAIVNIVKIP--PLOADKEDPISKSEIVVOFPCSDRFKPSYHSFST 246
 242 ENYIVFIEDPFIKIDYKLTAY-IRGVNNASCLSFHKEDKTMFHFVDRKTRKSTKFT 300
 247 PNIVIVETPVKINLFKFLSSMSLMGANYMCFESNELTMGVALIADKKRKYINNKRT 306
 301 DALVIVHINAYEEDGHVVDIAVARD--NSLYDMFYKLLDKDF--VNNKLISITLCK 356
 307 SFENLFRHINTVEEHEFLVDLQCKJFFVYNYLYANLRFNMHFFVKKNAPKAPQEVF 366
 357 KRVNPLINIK-ADTKNLTATTPNTATATLSTHTIMEFPVIFSGPQAFEPFQINQK 425
 411 YNGKRYVYATEVQSPVETRIAKLNQSTKEVILMGSTHWPSEFIVPSEFLAEEDES 476
 426 YGKRKYTVAYIGLNIHF-VPDKLKIKINATKRETWMQEDRSYSPRTIVSHPTALEDR 484
 471 VITVIVVSFPKAP-FITIDAKTFKFGATVAVENHIGI HSNF 515
 485 VLVSVVSPGAGQKPAYILILNAKDLSEVAPAFVEINIPVTFHDLF 530

QY 244 YIVFIEQPK:DIYKLTATV INGVNMAAGL SFHKEDK TWFEVDRKTKKEVSTKFT 300
 DB 248 YIVFIEVPRKINIKRELISWLSWGANMDC FESFTNETMGIWHLIDKRRKRYLNKKYPT 297
 QY 301 DALVLYHHINAAYEDGVNFDIYAARD-NSLYDMF--YLKRLDKDE---VNNKLSIPT 354
 DB 298 SFPMFLHHINTYEENGLIYDLCOMKGEFVYNYFTLYLANLENNEEVKKAKRKPQPE 357
 QY 355 CKRPVPLQYDKDAEVGSNLVKLP-TSATAV--KEKDGSTYQPELCEG---IELPV 407
 DB 358 VRRVVLPLINDK-ADTEKNLVTLPTNTATALCSDDEFTTWLEPEVLFGSPROAFEPDPI 416
 QY 408 NYD-YNGKKKKYVVAIEVQMSPVPTKIAKLVQIKVELH--WGEDHCEPSEPIFVPSIPA 464
 DB 417 NYQYCGKRPYTAAGDLNHF-VPRDLCKLVNKTREKLTWETVMOEPDPSSEPIFVSHPA 475
 QY 465 REEDGVVLTGVVSEPNKAP-FLIIDAKTFKELGRA--TVNVEMLDLHGKF 515
 DB 476 LEEDGVVLSVSVSGAGCKPAVILLINAKDLSEVAKAEFTVEINIPVTFHGLF 529

RESULT 14

ABH71747 standard; Protein: 620 AA.

ABH71747:

26-MAR-2002 (first entry)

Drosophila melanogaster polyprotein; SEQ ID NO 42033

Drosophila, developmental biology; cell signalling; insecticide;

pharmaceutical.

Drosophila melanogaster.

W0200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-0509231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI: 2001-656860/75.

N-PSDB; ABL15850.

New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell

interactions -

Disclosures: SEQ ID NO 42033, 21pp + Sequence Listing, English.

The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is

useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of

insecticides, therapeutics and pharmaceutical drugs. The invention

discloses genomic DNA sequences (ABL1617b-ABL30511), expressed DNA

sequences (ABH57737-ABH72072).

The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from WIPO

at http://wipo.int/pub/publ/usho2_pat_sequences

Query Match 24.88; Score 849.5; DB 22; Length 620;
 Best Local Similarity 34.68; Pred. No. 420-67;
 Matches 203; Conservative 99; Mismatches 189; Indels 95; Gaps 18;

QY 9 KEHPPEPKAEVQSGQLPTWGLVLLPQNGMHTIDPKINIMHTLALHSPTPKNCFVY 68
 DB 50 FRFIVPTEBHSHSHPIKMYICSLIDRNGQSNWVDMTQGHLPVSSALDHRAPIANCVY 109
 QY 69 YKSTYLSISQTYN-NIENKLVSEFGLMAYFQKNIKPAKASTYLSHTTPE TQWLLN 127
 DB 110 YQNPFDVETELKPKRSQRIIVTEFPIAAYDPCHSIFDY---FAAFEPNPGTINMSLS 166
 QY 128 IMKGDYVATSETNFIKIDPOTLETLDKVDKVAANLATSHPHYASGNIIMKTS 187
 DB 167 IYFGDDYVTFTEFPHRINPCTLATEARLCTTDPGVAVNHTSHHVLPSSTVNLCT 226
 QY 188 LVDKGRTKYLFLPSSVPEKPKSCFKHLEVCSIPKSLQSYNSPGLIENYVYF 247
 DB 227 MTPSGPAYTLL-----SFHSGDY---MFEIAHVAVTLPTKMKIIPDYMTPTDIDHVPV 278
 QY 248 IEPPEFLDIYELATAYIPGVNMAAGLSEFHEKRTWFEVDRKTKKEVSTKFTYALVYH 407
 DB 279 VEPPLSVSTLEYTKAOLAGQNLNALKMFEHMLFLHLIDKRSKLVQI YSEKAPVYH 457
 QY 308 HINAVFTGSEVTVVAVKRIETVMVYPTVKKFTVYVKKPTTCTTCTP PVVYH 462
 DB 338 IINCERDGHVVVVICSYKRPENMINMYLFAI-ANNQININVAITLFRGHLIFVLTGII 496
 QY 363
 DB 367 PFIASIANPDLVKSSTLAGISAVGVSLIPKNSVSNALVYQWQVQALVATLHESKPKQK 456
 QY 370 VG-----NLVKKLP-----SALAVEKELSGSYVPELEDE-LELPRVNVY-YNKPKRYVY 419
 DB 457 RGRYEENLVNIVTMESSQAELAQNG-IGLRPEMLCOMGCEPRTIYERHMKKNYVYF 515
 QY 420 YATVQMSPV-PTKIAKLVNQTREVLHMGDHCWSEPIFVSPAREHDSKVLTGVVY 478
 DB 516 YAISSDVAVNPGLIIVDVWKNKSLTWCSEENVYSEPIFVSPPKSKSDGVILLASVYL 575
 QY 479 SEPN-KAPFLIIDAKTFKELGPAIVNVEMLD-----LHGMPFIP 517
 DB 576 GGLNDKRYVGLIYLAKMTITLHGK-----DFHTNIPVYKRLIPMPVAF 617

RESULT 14

AAU04291 standard; Protein: 620 AA.

AAU04291:

23-OCT-2001 (first entry)

Drosophila beta-carotene dioxygenase (beta-diox 1).

Beta-carotene dioxygenase; beta-diox 1; Drosophila; beta-carotene;

lycopene; beta-apocarotenal; beta-ionone; apolytrypenol; grain seed; corn;

oil seed; palm; chick pea; diagnostic; therapeutic; lycopene;

retinoid/vitamin A deficiency; beta-diox 1; transgenic; nutrition;

carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop;

Drosophila melanogaster.

W0200148153-A2.

05-JUL-2001.

27-DEC-2000; 2000WO-EP13273.

24-DEC-1999; 99EP-0125895.

20-MAR-2000; 2000EP-0105822.

(GREE-) GREENOVATION PFLANZENBIOTECHEMLOGIE GMBH.

XX Von Lintig J, Vogt K;
 PI MPI: 2001 425657/45.
 XX
 DR N-PSDB: AAS07180.
 XX
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
 PT (beta-diox II) protein that cleaves beta-carotene and lycopene to yield
 PT beta-apocarotenal and beta-ionone, and apolycopeneals, respectively
 XX
 PS Disclosure; Fig 6; 116pp; English.
 XX
 CC The sequence represents the amino acid sequence of Drosophila beta-
 CC carotene dioxygenase- (beta-diox II). Beta-diox II specifically cleaves beta-
 CC carotene and lycopene to form beta-apocarotenal and beta-ionone, and
 CC apolycopeneals, respectively. The DNA is useful for transforming grain
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm,
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes;
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or
 CC ribozyme type therapeutic agents and for detecting any abnormality of
 CC endogenous beta-diox II. The beta-diox II specific antisense
 CC oligonucleotides derived from the DNA sequence are useful for dose
 CC response studies in relevant models of retinoid/vitamin A deficiency
 CC during any stage of an organism's development. The nucleic acids are also
 CC useful as probes and as a guideline to define new PCR (polymerase chain
 CC reaction) primers for the cloning of substantially homologous DNA
 CC sequences from other sources. The nucleic acids are also useful for
 CC determining the presence or quantity of beta-diox II nucleic acid and
 CC useful for increasing or decreasing the amount of beta-diox II levels in
 CC a cell or tissue which can modify the level of vitamin A and other
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,
 CC screening of an expression library to identify nucleic acids encoding
 CC beta-diox II or the structure of functional domains. The transgene
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole
 CC organisms have improved nutritional quality or physiological condition
 CC and accumulate important metabolites of carotene/retinoid pathways such
 CC as vitamin A aldehyde and retinoic acid, beta-carotene or take up beta-
 CC carotene from the medium. Expression systems involving beta-diox II are
 CC useful in the study of beta-diox II activity. Identification of cDNAs
 CC encoding beta-diox I and II allows the physiological characterisation of
 CC mammalian, carotene/retinoid metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.
 XX
 SO Sequence 620 AA;
 Query Match 29.8%; Score 840.5; DB 22; Length 620;
 Best Local Similarity 34.5%; Prof. No 3 2e-67;
 Matches 203; Conservative 99; Mismatches 189; Indels 95; Gaps 18;
 OY 9 KEHPERIKAEVOCGLPTWGLVLRNPGMHTTDDTKYNNWFGCLALLHFTKNGEYV 68
 DB 50 EREIVDPFGHSHGHPKMGISLNRNPSKWKVLMFGSHLPPCSALLHFAALRNGAVT 109
 OY 69 YRSKYLRSDTYNCLNEANRIVSEGTMAYPDCKNIFAKAFSYLSPTTPEF-DNCLN 127
 DB 110 YNRFVDETLLKKNSSAGRIIVTRGTAAVPRPCISIFPR---FAATFRPISGTPNMSIS 166
 OY 128 IMKGGDYATSETFNFIKIDQLETLKDYNSKYVANLAFHFDYASGNTINMST 187
 DB 167 IPPFGQVYVTFETFPFMRINPTCTATFARICITTFVSVVNHSHPHVLPSTVYVNT 226
 OY 188 IYDKGRTKTVLFKISSVPEKKKSCFKHLEVOSIPSPSLQPSYHSGTENTIVE 247
 DB 227 MRPSPAVTTL---SPPTED---MFAHVAATFCTPFWLHGJYHFTGLTHFVI 278
 OY 248 IEOPEKLDIVKATAYIRGVNWSASLSEFKEDKTMFHVVDKRTKEVSTKYFTALVYH 307
 DB 279 VEOPLSVSLTEYIKAGLGQNLASLKFEDRPLFHLHLGVNSKIVQI-VKSHAPFYIH 347
 OY 308 HINATFEGSHVVEDIVATPNSLITPMFLIKKLKDFVNNKLSTIPCK--PRVAVI--- 362

DB 338 IINCEERHSHVVVDITSTNRPMTNMTLEAT-AMQINPVNAILLEKRLPKEVPLGIL 356
 OY 363 -----OY-----DKDAE 369
 DB 397 PRASIAKRGVKSFLSAGISAPQVSRIMKHSVSQADITTPYRNGKATAGENSPPKDAK 456
 OY 370 VG-----SNVKLPT-----SATAVKEKSSIEGPEELICE-GIELPPVNY--YVNHKAYV 419
 DB 457 RGRYEENLVNLTWEGSQAEAFQGTNG-IQLFREMIDMGCETPRIYERKMKRYK 515
 OY 420 YATVEQWSPV-PTTATKLVNVTTFVLIHMGCPHCPSPPTIYVSPDAPPPPVAVIIVVV 478
 DB 516 YAISSVVAVNDGTLIKVYVWKSCITWCFEYVSPPTIYVSPDAPPPPVAVIIVVV 575
 OY 479 SRPN-KAPLLIIDAKTFKELGRATVYEMHLD-----LHGMPIP 517
 DB 576 GGINPQVVLIVLCAKMTLEUSFC---DPHTNGPVPCILHGWAP 617
 RESULT 15
 ID AAG62841 standard; Protein; 620 AA.
 AC AAG62841;
 XX
 DT 17-SEP-2001 (first entry)
 XX
 DE Amino acid sequence of a beta-carotene dioxygenase (beta-diox).
 XX
 KM Beta-carotene dioxygenase; beta-diox; beta-carotene; vitamin A aldehyde;
 KM transgenic plant.
 XX
 OS Drosophila melanogaster.
 XX
 IN W020014162-A2.
 PD 05-JUL-2001.
 XX
 IT 22 DEC 2000; 2000MO-RP13144.
 XX
 PP 24-SEP-1999; 99EP-0125895.
 PP 20 MAR 2000; 2000EP-018-822.
 XX
 PA (GREF-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.
 XX
 PI Von Lintig J, Vogt K;
 XX
 DR MPI, 2001 441713/47.
 DR N-PSDB: AAH42156.
 XX
 PT Dioxygenase (beta-diox) protein that cleaves beta-carotene to form
 PT vitamin A aldehyde, and polynucleotides encoding them useful for
 PT producing transgenic bacteria, fungi, plants expressing the polypeptide
 XX
 PS Claim 7; Fig 6, 87pp, English.
 XX
 CC The present sequence represents a beta-carotene dioxygenase (beta-diox)
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic
 CC plants. The transgenic plants have improved nutritive quality or
 CC physiological condition and accumulate vitamin A aldehyde and can take
 CC up beta-carotene from the medium. Expression systems encoding beta-diox
 CC are useful in the study of beta-diox activity. Identification of cDNAs
 CC encoding beta-diox allows the physiological characterisation of
 CC mammalian vitamin A metabolism. Vitamin A production in crops and
 CC microorganisms can be achieved by transforming the organisms or crops
 CC with the cDNAs.
 XX
 SO Sequence 620 AA;
 Query Match 29.8%; Score 840.5; DB 22; Length 620;

Best Local Similarity 34.68; Pred. No. 3,2e-67;
Matches 203; Conservative 99; Mismatches 189; Indels 95; Gaps 18;

```

QY      9 KEEHPEPIKAEVGOQLPTWLOGVLLRNGPGMHITIGTKYHMFQDGLALHSTFKNGEYV 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     50 EREIVDPLECHHSHIHKWICGSLRNGPGSMKVGDMTPGHLFDGCSALLHFAIRNGRYT 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     69 YRSKYLRSDTYNCNIFANRIYVSEFGTMAVPDPCKNIFAKAFSYLSHTIPEF-TDNCLIN 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    110 YQNKRFVDFTLRKKRKSQRIVYTFEGTAAPDPCHSIFDR---FAAIFRPSDSTGDSMIS 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    128 IMKTGDDYATSETNFIKKIDPQLETLLDKVDYSKYVAANLATSHPHYDSAGNILMGYS 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    167 IYPPGDDQYTFTEFEPFHMRIINPCTLATFAPICITDEGVYVNHSHPHVLPSGTIVYNLGTT 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    188 IVDGRTIKYVLFKIPSSVPEKEKSKCFKLEVVCSIPSKLLQPSYHSFGITENYIVF 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    227 MTRSGPAYTII-----SPPHGQ---MFEHAHVAVATLPCRMKILPGMHTEGLTDHYFVI 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    248 IEOPEKLDIVKIAVAYIKGVNMAQSLSFHEDKTFHFVDKRTKKKEVSTKFTYDALVLYH 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    279 VEOPIVSLSLTFEYTKAQIGSGNLSACLKMFEDPPTLFLIDRVSGKLVOI-YESEAFFYLH 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    408 HINAVEEGHVVPHIIVAYFQNSLYDMFYLLKLLKDEEVNKKLISIPCK--RFVPL--- 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    348 IINCFFRQGHVVVDICSYFNPDEMIMMYLFAI-ANMOTNPVATLFEQGPLREVLPLCTI 396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    363 -----QY-----DKDAE 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    397 PPASIAKRGIVKSFSLAGLSAPQVYSKIMAHVSQYADLIIMPIINCKVATAGEESPCKDAK 456
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    370 VG----SNLYKILPT---SATAVKEKDSIYCQPEILCE-GIELPVNYD-YNGKKRYV 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    457 KGRYEELNLVNLVTMEGSOAEAFQJING-IOLKPEMLCJMGCEPRIYERYWGNRYF 515
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    420 YATEVQMSPV-PTKIATLNVOTKEVLHMGEDHCPSPPIFVPSDAREDEGVVLTGVV 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    516 YAISSDVDAVNPCTLIKYDVWNKSCLTWCBEENVYPSSEPIFVPSDPKSEDDGVILASMYL 575
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    479 SEPN-KAPFLLILDAKTFKELGRATVVENMILD-----LHGMFIP 517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    576 GGLNDRYVGLIIVCAKIMTELGRG---DPIHTNGPVPKCLHGMFAP 617
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 15, 2003, 09:27:36
Job time : 75 secs


```

Db      307 SPENLFIHINTYEDHEFLIYDLCWKGEEFYVNSYLANLRKEMWEEKKAKARAPQREVR 366
Oy      357 KFYVPLQYLUKAEVGSNMLALF-1SALAWEKESUYCPELCEG-----TELPKVVND- 410
Db      367 KYVPLPLNDK-ADTGKNLVTLPLNTATATLCSDETLMLEDEVLFSGRVAFEPPOINYOK 425
Oy      411 YNGKKKKYVATEVQWSPVPTKLAKLVNOKTEVLMHGEDHCWSEPTFVPSPPAREPDRG 470
Db      426 YGCKPYTAYGLGHNH-VPDRICKLVNKTETWQFQPSYSEPTFVPSHPALBEDGG 484
Oy      471 VLTGVVSEPNKAP-FILLDAKTEKELGRATVNMMLDLHGMF 515
Db      485 VLVSVVSPGAGQKPAYLLINAKDLSEVARAEVEINIPVTFGLF 530

```

RESULT 2

```

P88115
Protein F53C3.12 [Imported] - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C.Accession: F88115
Anonymous: The C. elegans Sequencing Consortium
Science 282, 2012-2018, 1998
A.Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A.Reference number: A75000; MIM:906613; PMID:981916
A.Note: see websites genome.wustl.edu/gen/c-elegans/ and www.sanger.ac.uk/Projects/C-ele
A.Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A.Accession: F88115
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-556 <STO>
A.Cross-references: CB,Ch11, FICN A767462.1, P1103786489, GSPR-GRN0062, GSPR-F88115,
C.Genetics:
A.Gene: F53C3.12
A.Map position: 2

```

```

Query Match      27.6% Score 778.5; DB 2; Length 556;
Best Local Similarity 31.7% Pred. No. 1.2e-47;
Matches 163; Conservative 110; Mismatches 233; Indels 19; Gaps 11;

Oy      14 EPIKAEVQGLPTVLQGLVLRNGPMHTIGDTKYNMHFDLALLHSFTFKNGEVYRSKY 73
Db      44 EPKLSTSGSVPSLTKMLKNGMFEIDTKYHFDGMSFLQKHEEDGKMYSAK 103
Oy      74 LRSTYCNIEANRIVSEFGTMAYPDPCKNIFAKAFSYLSHTIPEFTDNLINIKTGD 133
Db      104 LSENKYKKMEAGRIYSGFGTASFDPCKSIFSRFSSVQS-EGIHNNAYAFAPVD 162
Oy      134 DYATSETNFIKIDPQTELELDKVDYSKYVAANLASHPHYDSAGNIIMGTSTVDKGR 193
Db      163 GLVACTETPNKHHPDLASHTTIEPVNPSKYVALHTCTAGCLPENCDDVYNSRGPDA 222
Oy      194 TKYVLFKIPSSVPEKEKKSCFHLLEVCSIPRSLLQPSYHSFGITENYIVFIEQPK 253
Db      223 HVTIVTKNPKNL-QSDSDRS-WEHTKIGELRCSETFEYPTMHSFMSENYILIMPESPIR 280
Oy      254 LDIVKATATIRGVNMASSCLSFHKEDKTHFVDRKTKKVSRTFTDALVLYHNINAYE 313
Db      281 IDIKKELMKRPITTTFFDCKMKHADQKIFILNKKTGEOVPLKLMAPEFTFHHANTE 340
Oy      314 EDGHNVEDIAYVRDNLSDYDMFYLLKLDKDFEVNKKLTSIPTCKREVVPLQYOKDAEVS 373
Db      341 KKGGLVVDYCRIDAGNFDALEENM-KTGNKQNDALFPLYLKVLILPLSLIYGAAPGDD 399
Oy      374 LVK--LPTSATAVKEKDSIVQPELICE-GIELPRVND-YNGKKKYVATEV--Q 425
Db      400 LKPLGMAKGSALFQDDGKIRLKEKRVFTDISMEFPYHMEKINMKPYVYVSSVLAD 459
Oy      426 WSPVPTKIATLNOTKRVLMGSDH---CWSEPTFVPSPPAREDECVLTCVYVSP 481
Db      460 KSEPLGVIVKADLENGDHKVPMPENDKQIT--GEPIFVNPESVPELDDGILLIVVMTSD 517

```

```

Oy      482 NKAPELLILAKIFKEDSPALV-NVEMHLLHGMF 515
Db      518 GQRPVILLEKKNLTETARTTIEARIPDGFHAFY 552

```

RESULT 3

```

hypochelatal protein T16118.20 - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse ear cress)
C.Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C.Accession: T10688
R.Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bat
submitted to the Protein Sequence Database, June 1999
A.Reference number: 216533
A.Accession: T10688
A.Molecule type: DNA
A.Residues: 1-616 <BEV>
A.Cross-references: EMBL:AL049915, GSPR-GRN0062, A753716.20
A.Experimental source: cultivar Columbia; BAC clone T16118
C.Genetics:
A.Gene: A1SP:T16118.20
A.Map position: 4
A.Introns: 105/3; 359/3; 425/2; 549/1

```

```

Query Match      15.5% Score 437; DB 2; Length 616;
Best Local Similarity 28.4% Pred. No. 3e-23;
Matches 143; Conservative 68; Mismatches 189; Indels 104; Gaps 21;

```

```

Oy      20 WQSDLEPLWLVLLKNGSLNMLLIDIKKINWHEMLALHSFTFKNGEVYRSKY 79
Db      97 VQGRTPVWLVNFKYKSNLGNIGDHPKHLFQYSFLVQLPFGGRIPAAHRLLESDAY 156
Oy      80 MCNLEANKIVSEFGTMAYPDPCKNIFAK-AEYSLSHTTPE-----TANGLINIKTGD 134
Db      157 KAACKHNPICYPERS-----ETPKSVIINKNPSGIGFIVPLFSGSLTNNANVGIKRLD 212
Oy      134 D-YTATSEINFIRK----IDVULELDKVDYSKYVAANLASH-HEHTYSAGNIIMGTST 187
Db      213 GRVCLTET--QKSLVGHETLETIGKFTYDDVLESDIMLSAHITET---FMWTL 265
Oy      188 IVDGRTKRYVLPKIPSSVPEKPKKSCFHLFVWCSIPRS-LLQPSYHSFGITENYIV 246
Db      256 IPELVKRYVNPVPEASNKP-----EVVQVYKCSLSMKGWHSIAVLENNYV 315
Oy      247 FIEQPKLIDIVKATATV--IRGVNMASSCLSFHKEDKTHFVDRKTKKVSRTFTDAL 303
Db      316 IPEMPLKYSKNNLLKAEPIPLKFEWC-----PQLDAP IHVMSKLI-GEVVAASVAPAY 368
Oy      304 VLYHNINAYEEDGH-----VVEDIAYVR-DNSLDYDMFYLLKLDKDFEVNKKLTSIPT 355
Db      369 VIFHFINAYEEDKQSKAIVIIADCCENHADRILIDMLRLDTL-----RSSHGR-VLIDDA 424
Oy      356 K-RFVVPLQYOKDAEVSNMLVRLPTSATAVKEKDSICQPELICEGIELPRVNDYNG 413
Db      425 FLSGERFLD-----GSKYCKLEIAVEN-EEKH-----FAMLMCSLMLVLYG 465
Oy      414 KKYKYVATEVQW-----SYVPTKIATLNOTKRVLM 446
Db      456 QKYKYVYVAGQAPKICNPNALSKVYIYGLIYLSVNLNPNVMPVAVIVKRVKRW 475
Oy      447 GEHHCWSEPTFVPSPPAREDEG 470
Db      526 HEHGMIPSEFPVPRGATHEHDS 549

```

RESULT 4

```

S76169
hypochelatal protein - Synechocystis sp. (strain PCC 6803)
C.Species: Synechocystis sp.
A.Variety: PCC 6803
C.Date: 26-Apr-1997 #sequence_revision 26-Apr-1997
C.Accession: S76169
R.Kaneko, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asanuma, F.; Nakamura, Y.; Miyajima,

```